AD

AD743187

AMSAA

TECHNICAL REPORT NO. 59

COMPARISON OF BAYESIAN AND CLASSICAL ANALYSIS FOR A CLASS OF DECISION PROBLEMS

Erwin M. Atzinger
Wilbert J. Brooks

April 1972

U.S. ARMY MATERIEL SYSTEMS ANALYSIS AGENCY
Aberdeen Proving Ground, Maryland

63

R

DISPOSITION

Destroy this report when no longer needed. Do not return it to the originator.

DISCLAIMER

The findings in this report are not to be construed as an official Department of the Army position.

WARNING

Information and data contained in this document are based on the input available at the time of preparation. The results may be subject to change and should not be construed as representing the AMC position unless so specified.

ACCESSION for	
NTIS	WHITE SECTION ☑
DDC	BUFF SECTION ☐
UNANNOUNCED	☐
JUSTIFICATION	
BY	
DISTRIBUTION/AVAILABILITY CODES	
DIST.	AVAIL. and/or SPECIAL
A	

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) US Army Materiel Systems Analysis Agency Aberdeen Proving Ground, Maryland	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE
COMPARISON OF BAYESIAN AND CLASSICAL ANALYSIS FOR A CLASS OF DECISION PROBLEMS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)
Erwin M. Atzinger
Wilbert J. Brooks

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
April 1972	64	7

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO. 1P765801MM11	Technical Report No. 59
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT
Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	US Army Materiel Command Washington, D. C.

13. ABSTRACT

This report is concerned with decision making under uncertainty for the class of problems where the uncertain parameter is the Bernoulli success probability, p. For decision making purposes the desired information is frequently the probability of meeting a specific requirement for p. This problem is analyzed from both the classical and Bayesian points of view. The use of the posterior beta distribution obtained from the Bayesian updating procedure is discussed for this class of decision problems. A method for constructing a prior distribution, and a detailed example of the updating procedure with emphasis on this method, are also presented.

A comparison is made of the Bayesian and the most popular classical point and interval estimation techniques. These techniques are not directly applicable in evaluating the chances of meeting a specific requirement for p. However, for certain non-trivial estimation problems, where a point of interval estimate is sufficient, the Bayesian procedure deserves consideration.

DD FORM 1 NOV 65 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

Unclassified
Security Classification

Unclassified
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Success probability Reliability Decision making Uncertainty Bayesian						

Unclassified
Security Classification

TECHNICAL REPORT NO. 59

COMPARISON OF BAYESIAN AND CLASSICAL ANALYSIS FOR A CLASS OF DECISION PROBLEMS

Erwin M. Atzinger
Wilbert J. Brooks

April 1972

RDT&E Project No. 1P765801MM11

U.S. ARMY MATERIEL SYSTEMS ANALYSIS AGENCY
ABERDEEN PROVING GROUND, MARYLAND

U.S. ARMY MATERIEL SYSTEMS ANALYSIS AGENCY

TECHNICAL REPORT NO. 59

EMAtzinger/WJBrooks/pal
Aberdeen Proving Ground, Md.
April 1972

COMPARISON OF BAYESIAN AND CLASSICAL ANALYSIS FOR A CLASS OF DECISION PROBLEMS

ABSTRACT

This report is concerned with decision making under uncertainty for the class of problems where the uncertain parameter is the Bernoulli success probability, p. For decision-making purposes the desired information is frequently the probability of meeting a specific requirement for p. This problem is analyzed from both the classical and Bayesian points of view. The use of the posterior beta distribution obtained from the Bayesian updating procedure is discussed for this class of decision problems. A method for constructing a prior distribution, and a detailed example of the updating procedure with emphasis on this method, are also presented.

A comparison is made of the Bayesian and the most popular classical point and interval estimation techniques. These techniques are not directly applicable in evaluating the chances of meeting a specific requirement for p. However, for certain non-trivial estimation problems, where a point of interval estimate is sufficient, the Bayesian procedure deserves consideration.

 Next page is blank.

CONTENTS

Next page is blank.

COMPARISON OF BAYESIAN AND CLASSICAL
ANALYSIS FOR A CLASS OF DECISION PROBLEMS

1. INTRODUCTION

In the materiel-acquisition, decision-making process, much of the critical decision information concerning a system's capability is provided through test and evaluation of the system. A test result can often be scored as a success or a failure resulting in a type of data classification characteristic of a Bernoulli process (i.e., a process in which there are two mutually exclusive possible outcomes on each trial and where the outcomes on any given trial or sequence of trials do not affect the outcomes on subsequent trials). For example, missile system test flight data that are scored only as a hit or miss can be placed in this category.

Historically, one of the major objectives in test and evaluation has been to estimate the unknown Bernoulli success parameter, p. Where a point estimate is sufficient, the maximum likelihood estimate (the observed proportion of success) is one of the most popular of the classical estimates.[1] On the other hand, if a measure of uncertainty is desired, the classical interval estimates, based theoretically on the binomial distribution or on some large sample approximation thereof, are often used.

It is recognized that other classical point and interval estimation techniques do exist which may be applicable to this class of problems: the moving average and exponential smoothing are two of the alternatives. For this report, however, only the previously mentioned point and interval estimation techniques are considered since they appear to be the most popular. Throughout the report these techniques will be referred to as the classical techniques.

[1] Mood, A.M.; Graybill, F.A.; Introduction to the Theory of Statistics, Second Edition; McGraw-Hill Book Company, New York; 1963; p. 178.

In most real world decision situations one is confronted with a requirement that the probability of success, p, exceeds some specified value. Using the classical interval estimate, the best that one can do is determine whether the required value of p lies within the interval. No probability statement can be made about the unknown true value of p lying within this interval, and certainly no statement can be made concerning the probability that p will exceed the specified requirement.

Since, for large complex systems, extensive testing can be prohibitively expensive in both time and cost, many decisions must be made with a limited amount of test data. Certainly, in such situations all available information should be taken into consideration. In particular, two sets of test observations frequently exist in systems test and evaluation; one is based on production hardware, and the other on non-production hardware (R&D and Industrial Prototype hardware). The population of interest is usually the production hardware, but certainly the non-production observations do contain some useful information. Given these two sets of observations, the classical analyst can either ignore the non-production data or consider the combined population. In either case, he still will not be able to assess the probability of meeting the requirement.

More recently, Bayesian procedures, tailor made to address this type of decision problem, have been applied. According to Bayesian philosophy, any quantity whose exact value is unknown can be treated as a random variable. Thus, a probability statement can be made as to whether such an unknown parameter does or does not lie in a calculated interval. In addition, a statement can be made concerning the probability of exceeding a specified level. Of equal significance is the fact that the Bayesian approach provides a mathematically tractable technique for combining prior information with objective test data.

The objective of this report is to critically examine the classical and Bayesian procedures in an attempt to expose to the reader the merits of the Bayesian approach for the Bernoulli parameter class

of decision problems described earlier. Although the Bayesian approach is not a panacea, it is not difficult to see the potential for this approach in light of the current emphasis on risk analysis and decision risk analysis in the materiel acquisition process.

Although certain topics presented in this report have been discussed by others, they are repeated here since, to the authors' knowledge, there does not exist a comprehensive investigation of the Bayesian procedure applied to this class of decision problems.

Specifically, for the Bernoulli process (success parameter p) problem, this report includes:

- A description of the standard Bayesian updating procedure.
- A comparison of the classical maximum likelihood and Bayesian point estimates with respect to expected squared error loss.
- A critical examination of the classical and Bayesian interval estimates.
- A discussion of the applications of the Bayesian procedure to the decision problem.
- A description of a proposed method for constructing a prior distribution from prior test observations and all other prior information.
- A detailed example illustrating the Bayesian updating procedure.

2. BAYESIAN UPDATING PROCEDURE

This section contains a detailed description of the Bayesian updating procedure. It is introduced here to familiarize the reader with the terminology and notation characteristic of the Bayesian approach. This information will aid the reader in understanding the topics discussed in the report.

In estimating the uncertainty in the estimate of the average success ratio (p) for any Bernoulli process, a Bayesian updating procedure can be used. For the class of problems considered in this paper, the observations can be logically broken into two classes: one is used to

estimate the prior distribution of p (ℓ successes in m trials) and the other is used to update this prior distribution of p (k successes in n trials). The conditional distribution of k successes in n trials, given p, is binomial and its probability density can be expressed as

$$f_{k|p}(k|p) = \binom{n}{k} p^k (1-p)^{n-k},$$

where $k = 0,\ldots,n$. If p is assumed to have a beta distribution with parameters ℓ and $m-\ell$, then the probability density of p for $0 < p < 1$ and where $C(\ell,m)$* is the normalizing constant is given by

$$f_p(p) = C(\ell,m)p^{\ell-1}(1-p)^{m-\ell-1}.$$

Given this prior distribution of p, it is now possible to update this distribution with the n observations by applying Bayes theorem:**

$$f_{p|k}(p|k) = \frac{\binom{n}{k}p^k(1-p)^{n-k}C(\ell,m)p^{\ell-1}(1-p)^{m-\ell-1}}{\int_0^1 \binom{n}{k}p^k(1-p)^{n-k}C(\ell,m)p^{\ell-1}(1-p)^{m-\ell-1}dp}.$$

Combining the terms in the numerator, canceling out the constant term $C(\ell,m)$, and performing the integration, we reduce the above equation to

$$f_{p|k}(p|k) = C(k+\ell,\ n+m)p^{k+\ell-1}(1-p)^{n-k+m-\ell-1},$$

which is again a beta distribution with parameters $k+\ell$ and $(n+m)-(k+\ell)$. The mean and variance of the posterior beta distribution are

---

*$C(\ell,m) = \dfrac{\Gamma(m)}{\Gamma(\ell)\Gamma(m-\ell)}$

**Bayes theorem states that

$$f_{p|k}(p|k) = \frac{f_{k,p}(k,p)}{f_k(k)} = \frac{f_{k|p}(k|p)f_p(p)}{\int_p f_{k|p}(k|p)f_p(p)dp}$$

$$\mu = \frac{k+\ell}{m+n}, \text{ and}$$

$$\sigma^2 = \frac{\mu(1-\mu)}{(n+m+1)}, \text{ respectively.}$$

It should be noted that the posterior mean is the weighted average of the prior and update success ratio.

In certain instances, the analyst may want to weight the prior distribution. This can be done by applying a weighting factor, w, to the parameters ℓ and m which results in a prior distribution for $0 < p < 1$ of the form

$$f_p(p) = C(w\ell, wm)p^{w\ell-1}(1-p)^{w(m-\ell)-1}.$$

The updating procedure is then applied to this weighted prior distribution as in the unweighted case. The significance of the weighting factor and a method for selecting this factor are presented in Section 5.

The rationale for selecting a beta prior and a discussion of the application of this updating procedure to decision problems are deferred to Section 4.

3. ESTIMATION

3.1 Introduction.

In estimation theory, the general method used to estimate an unknown population parameter Θ is to select a random sample from the population, and then use the information contained in this sample to determine a point or interval estimate, say $\hat{\Theta}$, of the parameter Θ. For the class of problems being considered in this paper, the population of interest is that defined by a point binomial distribution* with the probability of success, p, being the unknown population parameter. In the following sections the classical maximum likelihood and Bayesian

*A discrete random variable X is said to have a point binomial distribution if its probability density function is of the form

$$f(x;p) = \begin{cases} p^x(1-p)^{1-x} & x = 0,1;\ 0 \le p \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

point estimation procedures and the classical and Bayesian interval estimation procedures are examined. The discussion is tailored to the class of problems being considered herein.

3.2 Point Estimation.

In point estimation, the squared error loss function,* $(\hat{\Theta}-\Theta)^2$, is often used to reflect the loss incurred in using the point estimator $\hat{\Theta}$ to estimate the parameter Θ. Since the value of $\hat{\Theta}$ is dependent on the sample data, $\hat{\Theta}$ is a random variable and the squared error loss is considered to be a random function of the parameter Θ. To eliminate the dependence on the particular random sample which is chosen, the mathematical expectation of the loss function, known as the risk function,** is used as an indicator of the quality of the estimator $\hat{\Theta}$. For squared error loss, the risk or expected loss reduces to

$$E[(\hat{\Theta}-\Theta)^2] = \text{Var}\ (\hat{\Theta}) + [E(\hat{\Theta})-\Theta]^2.$$

Thus, risk is a function of the unknown parameter Θ and is equal to the variance of the estimator plus the square of its bias.*** A good estimator is interpreted as one which minimizes the risk or expected loss over the critical range of the parameter Θ.

The classical maximum likelihood estimate of the population proportion, p, of a Bernoulli process generating n sample observations, is the observed sample proportion of successes, $\hat{p} = \frac{k}{n}$, where k is the number of successes and n is the number of sample observations. It has been shown that this maximum likelihood estimate enjoys many desirable characteristics. Among these are unbiasedness and minimum variance. Thus, within the class of unbiased estimates $\hat{p}$ minimizes the risk function or expected loss over the entire range of the parameter p.

---

*No justification will be given in this paper for the use of a squared error loss function. The interested reader should refer to Reference 1.

**Risk as defined here is often referred to as expected mean square error.

***Bias is defined as the difference between the unknown parameter and the mathematical expectation of its estimator.

The risk in this unbiased case reduces to the variance of $\hat{p}$ (i.e., Var $(\hat{p}) = \frac{p(1-p)}{n}$).* For these reasons, the maximum likelihood estimate $\hat{p}$ is often used, without reservation, as the best estimate of the population proportion p.

A fact which is often overlooked, however, is that a biased estimate is not necessarily to be rejected as inferior. The fact is that biased estimators exist which may, for a non-trivial range of the parameter p, result in a lower expected loss than the maximum likelihood estimate $\hat{p}$. Consider, for example, the Bayes estimator (Reference 1) $\hat{\hat{p}} = \frac{k+1}{n+2}$, which is derived using a squared error loss function and the assumption of a uniform or rectangular prior distribution for p. Since

$$E[\hat{\hat{p}}] = E[\tfrac{k+1}{n+2}] = \frac{np+1}{n+2} \text{ and } \mathrm{Var}\ (\hat{\hat{p}}) = \mathrm{Var}\ (\tfrac{k+1}{n+2}) = \frac{np(1-p)}{(n+2)^2},$$

the expected mean square error for the Bayes estimator is given by

$$\begin{aligned} E[(p-\hat{\hat{p}})^2] &= \mathrm{Var}\ (\hat{\hat{p}}) - (E[\hat{\hat{p}}]-p)^2 \\ &= \frac{np(1-p)}{(n+2)^2} - (\frac{np+1}{n+2})-p)^2 \\ &= \frac{1}{(n+2)^2}\ [(n-4)p(1-p) + 1]. \end{aligned}$$

Figures 3.1 and 3.2 exhibit risk as a function of the population parameter, p, for both the classical maximum likelihood estimate and the Bayes estimate for sample sizes of n = 10 and n = 100, respectively. Note that, for large sample sizes, they are approximately equal. However, for small sample sizes, which are often the case in many real world problems such as missile test and evaluation, the two can differ significantly. In particular, for n = 10, over the range $0.13 \leq p \leq 0.86$, the biased Bayes estimator is the better estimator in the expected squared error loss sense.

The curves in Figure 3.1 and 3.2 which represent the risk of the Bayesian estimator, were generated assuming a uniform prior

*Note that for the Bernoulli process described above, the random variable k has a binomial distribution with parameters n and p, Reference 1, page 182.

[1]Loc. Cit.

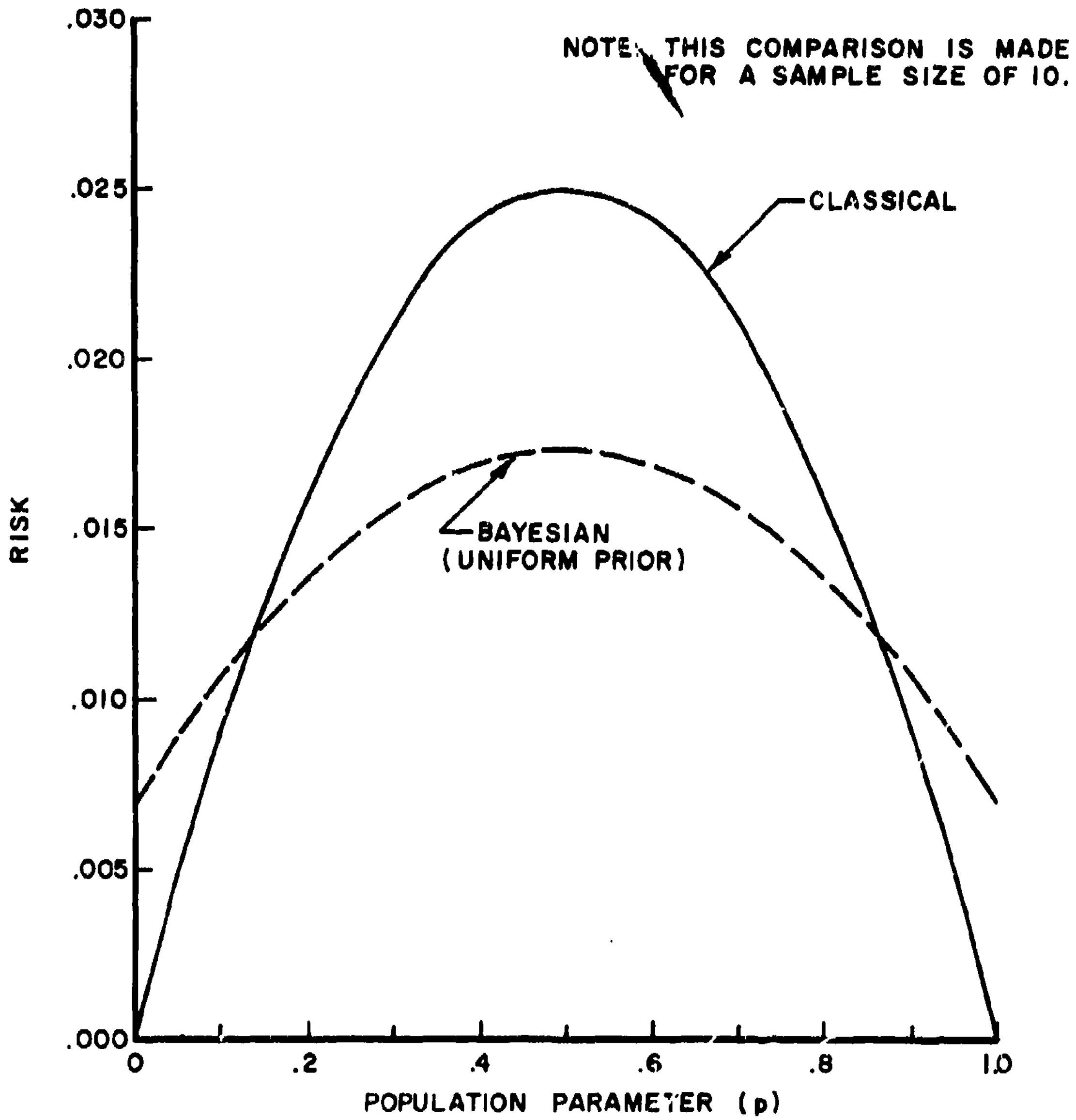

Figure 3.1 Risk* Comparison.

*RISK IS DEFINED AS THE EXPECTED SQUARED ERROR LOSS.

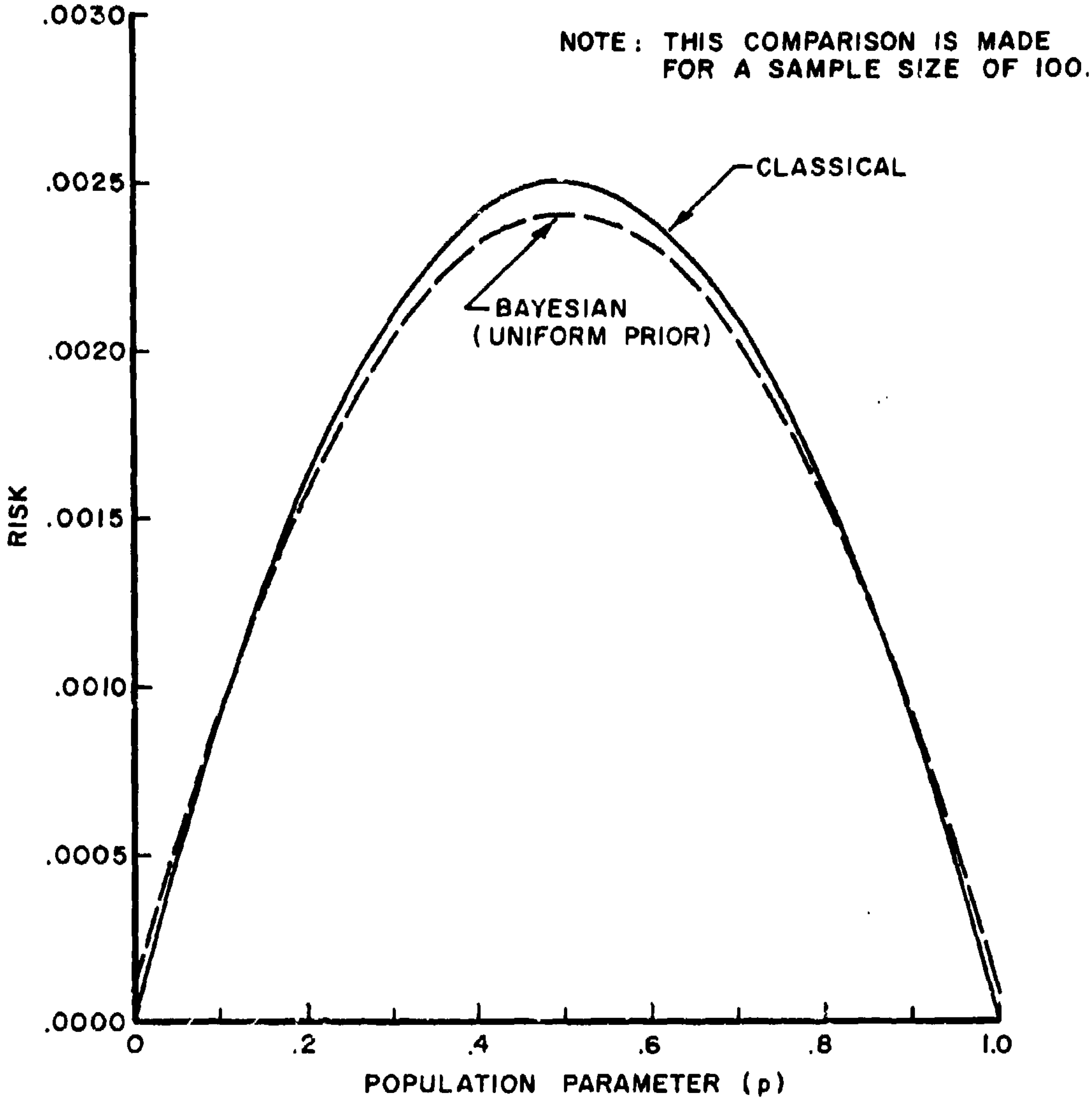

Figure 3.2 Risk* Comparison.

* RISK IS DEFINED AS THE EXPECTED SQUARED ERROR LOSS.

distribution. This assumption implies essentially complete ignorance on the part of the decision maker, of the value of the parameter being estimated. That is, all values of the parameter are assumed to be equally probable. This assumption is not only conservative, but is somewhat unrealistic for the class of problems under study in this paper.

To examine the impact of a more realistic prior distribution on risk (as a function of the population parameter p) the following comparisons are made:

1. A comparison is presented in Figure 3.3 of the risk for the classical maximum likelihood estimate for a sample size of 10, the Bayes estimate assuming a uniform prior distribution with 10 update observations, and the Bayes estimate assuming a beta prior distribution with parameters $\ell=3$ and $m-\ell=3$ (i.e., success proportion $= \frac{\ell}{m} = 0.5$) with 10 update observations.

2. A comparison is presented in Figure 3.4 of the risk for the classical maximum likelihood estimate for a sample size of 10, the Bayes estimate assuming a uniform prior distribution with 10 update observations, and the Bayes estimate assuming a beta prior with parameters $\ell=5$ and $m-\ell=1$ (i.e., success proportion of 5/6) with 10 update observations.

As can be seen in Figure 3.4, making stronger prior assumptions can be beneficial in some instances and detrimental in others. If the true population success proportion is less than 0.5, the risk in using the Bayes estimator with beta prior ($\ell=5$, $m-\ell=1$) is considerably larger than the risk of using either of the other estimates. On the other hand, if the true population success proportion, p, is larger than 0.5, the risk in using the Bayes estimator with prior beta ($\ell=5$, $m-\ell=1$) is substantially less than the risk associated with either the classical maximum likelihood or the Bayes (uniform prior) estimates. A decision maker who strongly suspects a population success proportion larger than 0.5 should, in the interest of minimizing his risk over the

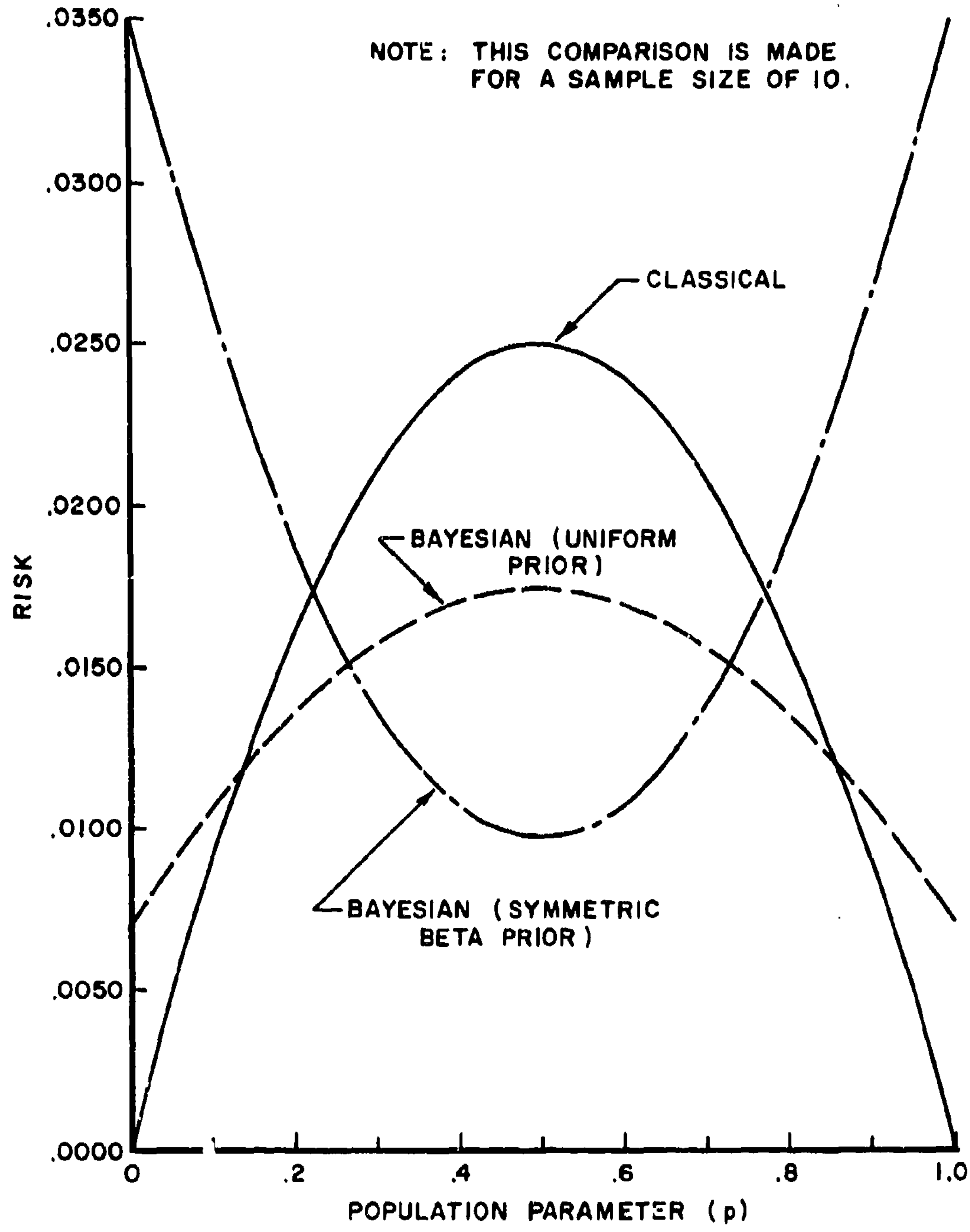

Figure 3.3 Risk* Comparison.

* RISK IS DEFINED AS THE EXPECTED SQUARED ERROR LOSS.

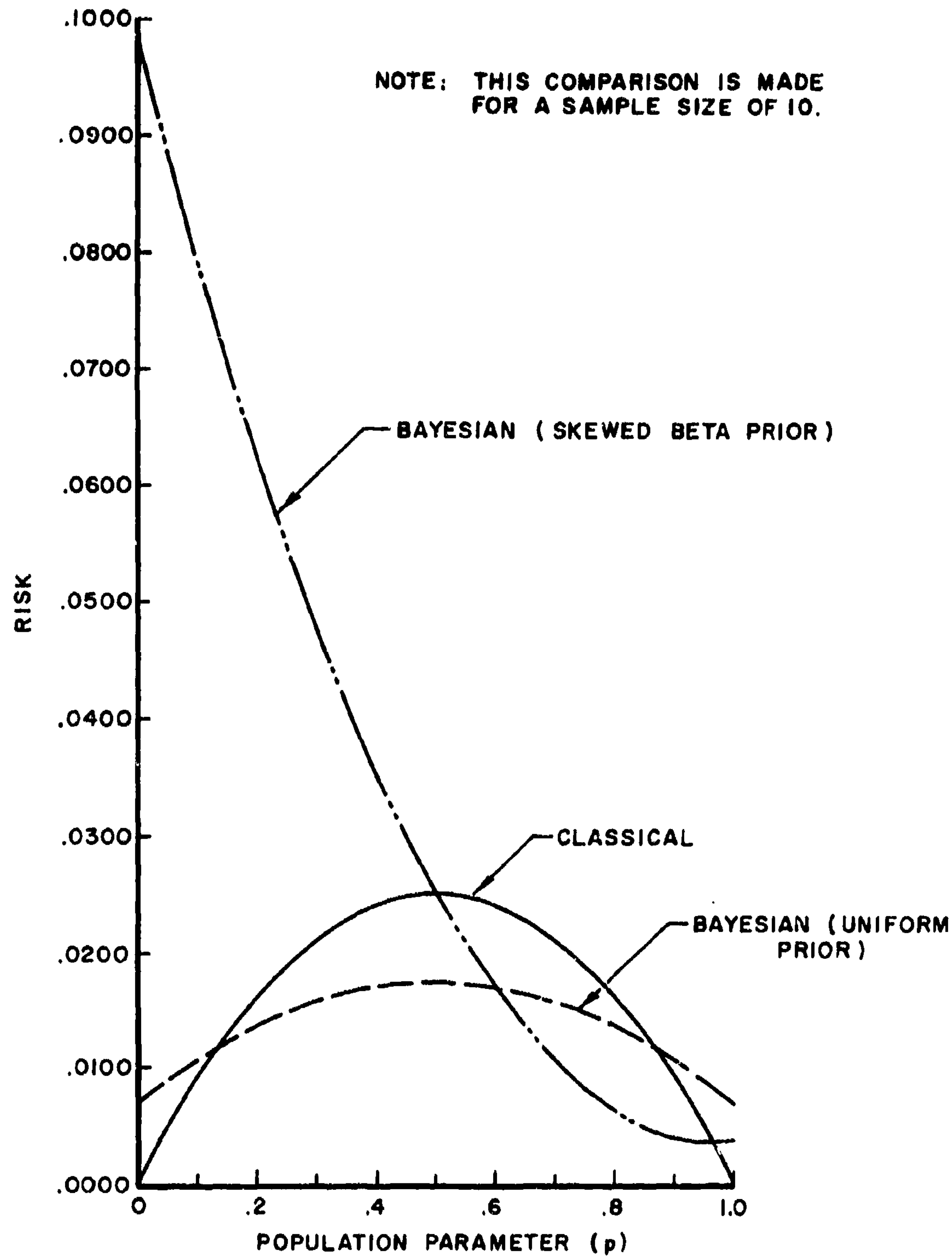

Figure 3.4 Risk* Comparison.

* RISK IS DEFINED AS THE EXPECTED SQUARED ERROR LOSS.

realistic range of this proportion, choose to use the Bayes estimate based on the beta prior distribution with parameters $\ell=5$ and $m-\ell=1$.

3.3 Interval Estimation.

3.3.1 General Description.

Generally, the single point estimate of the population proportion of successes, p, will be incorrect since the probability is very small that the estimate is exactly equal to the true population proportion. Some measure is needed of the uncertainty or error introduced in using this point estimate. This is certainly the case in risk analysis where the objective is to analyze uncertainty.

The classical statistical procedure most commonly used to account for this uncertainty is interval estimation. This procedure entails taking a random sample of n Bernoulli trials and then based on this sample, computing lower and upper confidence limits p_L and p_U, respectively. Associated with the confidence limits is a confidence coefficient, $1-\alpha$. The confidence coefficient is often misinterpreted as the probability that the true population proportion, p, will lie in the calculated interval (p_L, p_U). This would imply that the parameter, p, is a random variable contrary to classical assumptions. Actually, the interval, being the random variable in this approach, may or may not encompass the true value of the parameter, depending on the particular sample selected. What the confidence coefficient does represent is the proportion of such intervals which would be expected to cover the true population proportion if a large number of intervals were computed using independent random samples and the same estimation procedure. Of course, in most instances, the analyst cannot afford either the dollars or the time required to take repeated random samples. Thus, in practice, the analyst will act as though the interval is correct if the confidence coefficient is high. That is, for a confidence coefficient equal to 0.95, the analyst knows that the particular interval obtained from the sample data was generated by a procedure which would yield an interval that covers the population proportion, p,

for 95 percent of the random samples selected. It is for this reason only that he is willing to assume that the particular interval he generated does cover the population proportion.

The Bayesian approach to interval estimation dictates that the information contained in the random sample should be incorporated with prior information through the use of Bayes theorem (see Section 2). The Bayesian analyst contends that the decision maker is not interested in some specified proportion of valid estimates in the long run, but is interested in combining sample t ata with any prior information to make a correct decision.

In the Bayesian procedure, the population parameter, p, is assumed to be a random variable with a specific prior probability distribution. This revision is accomplished mathematically by using Bayes' theorem, and the result is called the posterior probability distribution of p. (Details of this procedure are contained in Section 2.) Lower and upper limits, p_L and p_U, can then be obtained such that the probability that p lies in the interval (p_L, p_U) is equal to some specified confidence level, $1-\alpha$. Thus, the popular confidence interval interpretation, which is incorrect in the classical case, is valid in the Bayesian framework. The following section defines and discusses classical and Bayesian confidence intervals in relation to the problem of estimating the proportion of successes in a sequence of n Bernoulli trials. Without loss of generality, the upper confidence limit is assumed to be equal to 1 and the discussion is limited to the lower confidence limit, p_L.

3.3.2 Definition of Classical and Bayesian Confidence Intervals. Given an estimate $\hat{p} = \frac{k}{n}$ of the parameter p of a Bernoulli process, the classical $100(1-\alpha)$ percent lower confidence limit p_L is defined as the value of p such that

$$\sum_{y=k}^{n} \binom{n}{y} p^y (1-p)^{n-y} = \alpha \text{ (Reference 1).}$$

[1] Loc. Cit.

Tables of the cumulative binomial distribution can be used directly to obtain a solution to this equation. Several alternatives are available, however, which simplify the computation considerably. The first is achieved by noting that the cumulative binomial is related to the incomplete beta function by the following relationship:

$$\sum_{y=k}^{n} \binom{n}{y} p^y (1-p)^{n-y} = I_p (k, n-k+1),$$

where

$$I_p(k, n-k+1) = \frac{\int_o^p u^{k-1} (1-u)^{n-k} du}{\int_o^1 u^{k-1} (1-u)^{n-k} du} . \qquad (1)$$

This function has been tabulated by Pearson[2] and is easier to use than the cumulative binomial. The lower confidence limit, in this case, is given by the solution to $I_p (k, n-k+1) = \alpha$.

A second alternative is obtained by recognizing that the expression on the right hand side of Equation (1) is $P [0 < U < p]$ where U is a random variable having a beta distribution with parameters k and n-k+1. Thus, the lower confidence limit, p_L, is given by the solution to $P [0 < U < p] = \alpha$ or $P [p < U < 1] = 1-\alpha$, where U has a beta distribution with parameters k and n-k+1.

In the Appendix it is shown that by applying the transformation

$$U = \frac{\frac{k}{n-k+1} V}{1+(\frac{k}{n-k+1})V},$$

the lower limit p_L reduces, in this case, to

$$p_L = \frac{1}{1+(\frac{n-k+1}{k})v'_{1-\alpha}} \qquad (2)$$

[2]Pearson, K. Ed., Tables of the Incomplete Beta - Function, University Press, Cambridge, England, 1934.

where $v'_{1-\alpha}$ is the 100(1-α) percent point of the F distribution with 2(n-k+1) and 2k degrees of freedom. Since tables of the F distribution are generally more available than those of the cumulative binomial distribution or the incomplete beta function, this alternative is clearly of practical value.

Note that the classical confidence limits discussed to this point are exact. Several methods do exist which, for restricted ranges of the parameters n and p, give fairly accurate approximations. However, for the sample sizes of the class of problems being considered in this paper, these approximations are generally inadequate.

The Bayesian lower 100(1-α) percent confidence limit, p_L, of the Bernoulli parameter p is defined as the solution to the equation

$$P[p_L \leq p \leq 1] = \int_{p_L}^{1} f_{p|k}\,(p|k)\,dp = 1-\alpha,$$

where $f_{p|k}\,(p|k)$ is the posterior beta distribution with parameters $k+\ell$ and $(n+m)-(k+\ell)$ which was derived in Section 2. Note that ℓ and $m-\ell$ are the parameters of the beta prior distribution which was used in the derivation. Thus, p_L is given by the solution to the equation

$$P[p_L < U < 1] = 1-\alpha,$$

where U has a beta distribution with parameters $k+\ell$ and $(n+m)-(k+\ell)$.

As in the classical case, the Bayesian lower confidence limit can be expressed in terms of the F distribution. By using the transformation

$$U = \frac{\frac{k+\ell}{(n+m)-(k+\ell)}\,V}{1+\frac{k+\ell}{(n+m)-(k+\ell)}\,V},$$

along with the theory introduced in the Appendix, it follows that

$$p_L = \frac{1}{1+\frac{(n+m)-(k+\ell)}{k+\ell} v'_{1-\alpha}}, \qquad (3)$$

where $v'_{1-\alpha}$ is the $100(1-\alpha)$ percent point of the F distribution with $2[(n+m)-(k+\ell)]$ and $2[k+\ell]$ degrees of freedom.

3.3.3 Comparison of Bayesian and Classical 95 Percent Lower Confidence Limit. In Figure 3.5, a comparison is made of the classical lower 95 percent confidence limit with the lower limits of three Bayesian procedures, each unique in its prior assumptions. A sample of n=16 update observations is used in the comparison and the results are presented as a function of the number of successes, k, which could result in the 16 trials. For the convenience of the reader, the results are also presented in tabular form in Table 3.1. Note that for $k < 12$ all three Bayesian procedures result in a shorter* confidence interval than the classical procedure. Even the most conservative Bayesian procedure, that corresponding to a uniform prior distribution ($\ell=1$, m=2), is better than the classical procedure over the entire range of k. This fact is true, independent of n and k, since the ratio of the Bayesian lower limit with uniform prior ($\ell=1$, m=2), given in Equation (2), to the classical lower limit, given in Equation (1), is greater than one.

In Figure 3.5, one should observe that, over a limited range of k, the classical procedure does result in a shorter interval than the Bayesian procedure with symmetric prior. This should serve to caution the user that the Bayesian procedure can lead to poor results if insufficient effort is devoted to the rational selection of the prior probability distribution. No one can argue the fact that the assignment of prior probabilities is a problem area in Bayesian analysis. Because the selection of the prior probability distribution involves using subjective judgment, many classical statisticians choose to rule out the Bayesian procedure as a viable estimation technique. The classicists choose to be somewhat conservative in their estimates and

*In using an interval to estimate a parameter, minimum length is a desirable characteristic of the estimator.

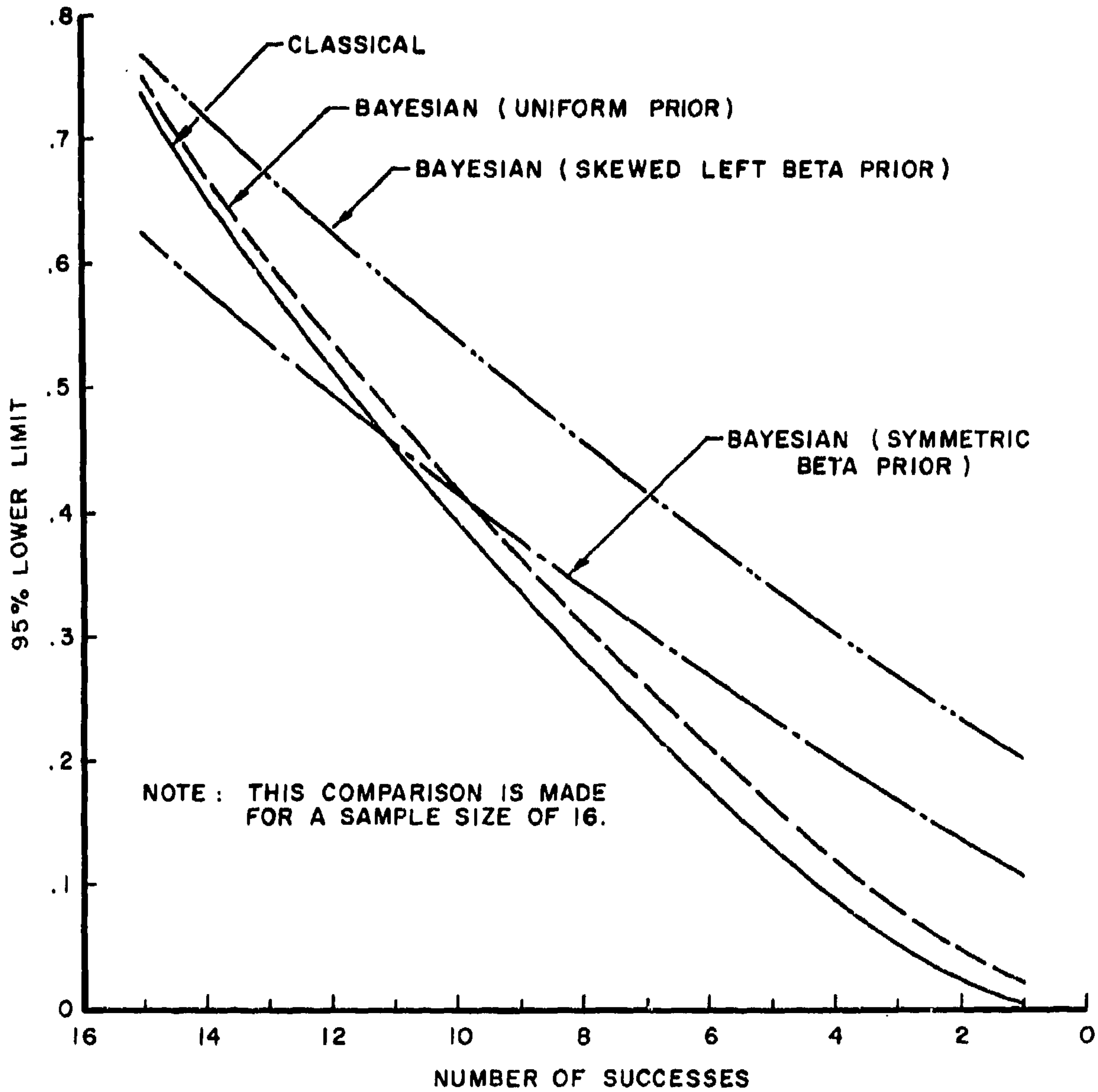

Figure 3.5 Comparison of the Classical and Bayesian 95 % Lower Limits.

TABLE 3.1 COMPARISON OF CLASSICAL AND BAYESIAN 95 PERCENT LOWER CONFIDENCE LIMITS (SAMPLE SIZE, 16)

No. of Successes	Lower Limit			
	Classical	Bayesian (Uniform)	Bayesian (Prior 4 out of 8)	Bayesian (Prior 7 out of 8)
15	.736	.750	.625	.769
14	.656	.674	.581	.718
13	.583	.604	.538	.671
12	.516	.539	.496	.625
11	.452	.478	.456	.581
10	.391	.419	.417	.538
9	.333	.364	.379	.496
8	.279	.311	.341	.456
7	.227	.260	.304	.417
6	.178	.212	.270	.379
5	.132	.167	.235	.341
4	.090	.124	.202	.304
3	.053	.085	.170	.270
2	.023	.050	.139	.235
1	.003	.021	.110	.202

make absolutely no prior assumptions concerning the parameter. It is the contention of the authors that, in today's decision-making world, the complete absence of pertinent information prior to sampling is rare. As the result of the test and evaluation of a system, some test results exist (R&D and/or Industrial Prototype test results) and/or engineering judgment upon which to base a prior distribution. Therefore, in many realistic problem areas the conservative classical lower limit can and should be improved upon.

The Bayesian procedure provides a technique for updating prior knowledge with sample data and permits one to be conservative, but does not force strict conservatism on the analyst as do the classical techniques. In later sections, further discussion is provided concerning the value of Bayesian techniques in decision-oriented problems and the rational selection of prior probability distributions.

The reader interested in examining the comparison between classical and Bayesian lower limits for other prior assumptions and other confidence levels is referred to a report by Benton.[3] In that report, tables are also provided for the 0.99, 0.975 and 0.95 Bayesian lower confidence limits (assuming a uniform prior) for sample sizes up to n=25.

4. THE BAYESIAN PROCEDURE APPLIED TO DECISION MAKING

In the Bayesian update procedure, presented in Section 2, a posterior beta distribution was derived by updating a beta prior distribution with Bernoulli type test data. A posterior beta distribution was seen, in Section 3, to be the driving force in both the Bayesian point and interval estimation of the Bernoulli success parameter, p. However, as discussed in the introduction, neither classical nor Bayesian point and interval estimation procedures directly address the decision

[3] Benton, Alan W., An Investigation of the Characteristics of Bayesian Confidence Intervals for Attribute Data; Technical Memorandum No. 14, November 1969; Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland.

maker's problem. Thus, it is in this decision making context that the Bayesian procedure, specifically the beta posterior distribution, proves to be most useful. Statistical interpretations based on the posterior beta distribution are much more realistic for the decision maker than any interpretations available using either classical or Bayesian point and interval estimation theory.

Before we proceed with a discussion of the advantages and disadvantages of the Bayesian procedure as related to decision making, recall that the posterior probability density of the parameter p (Bernoulli success probability) for $0 < p < 1$, is of the form

$$f_{p|k}(p|k) = \frac{\Gamma(n+m)}{\Gamma(k+\ell)\Gamma[(n+m)-(k+\ell)]} p^{k+\ell-1}(1-p)^{[(n+m)-(k+\ell)]-1},$$

where

n = number of update observations,
k = number of update successes,
m = number of prior observations, and
ℓ = number of prior successes.

Thus, p is a random variable having a beta distribution with $k+\ell$ and $(n+m)-(k+\ell)$ degrees of freedom. In Figure 4.1 the beta posterior probability density function is displayed for n=20, k=15, m=10, and ℓ=6. Its corresponding cumulative distribution function is provided in Figure 4.2. This specific case will be used as an illustration in the discussion to follow.

Faced with a decision concerning an uncertain parameter p, and given its posterior beta distribution, the decision maker has several options available to him. He can use the cumulative distribution function of the variable p directly to address questions such as the following:

- What is the probability of meeting a specific requirement for p?
- What is a more reasonable requirement if the above probability is unsatisfactory?

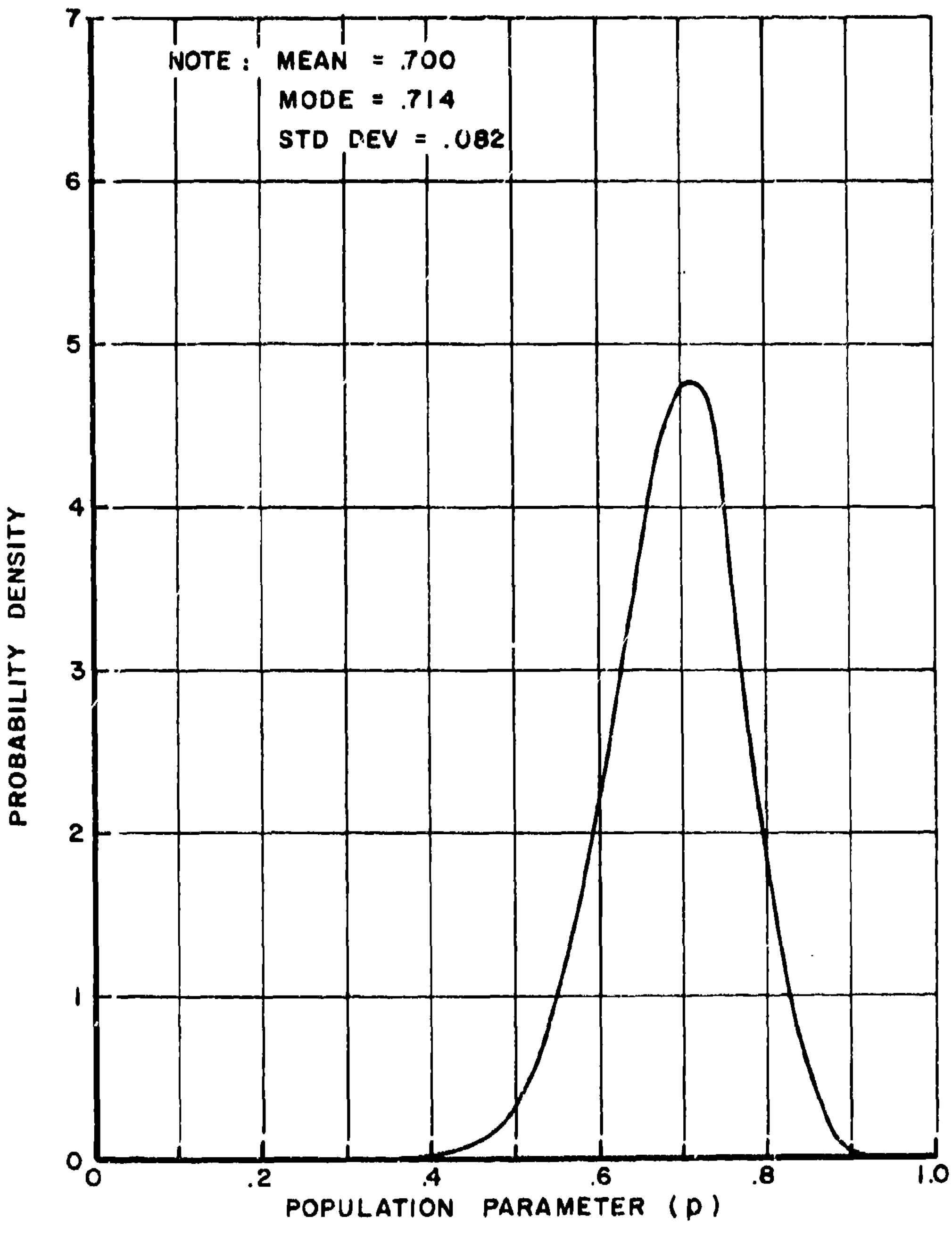

Figure 4.1 Beta Posterior Probability Density Function.

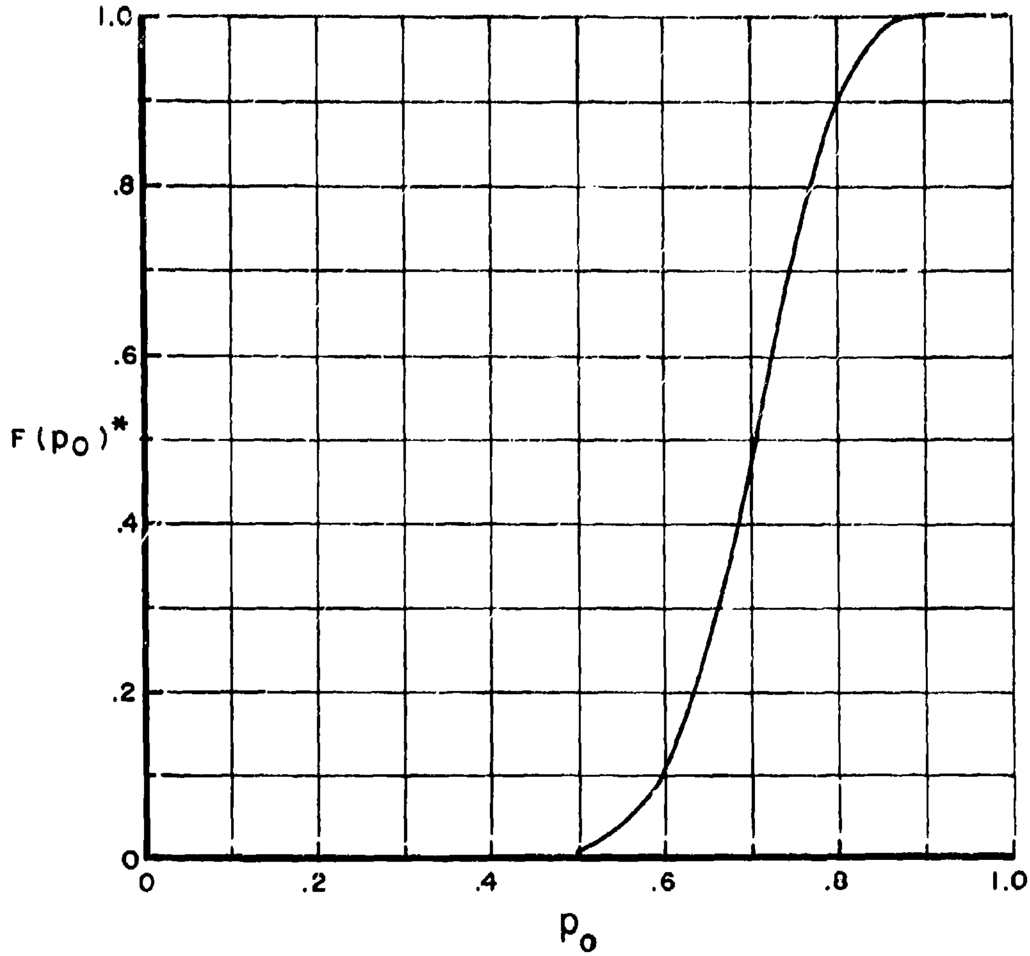

Figure 4.2 Cumulative Posterior Beta Distribution.

*F $(p_O) = P\left[p \leq p_O\right]$ (I.e. THE PROBABILITY THAT THE TRUE PROPORTION OF SUCCESSES (p) IS LESS THAN OR EQUAL TO p_O.)

For example, it is not too difficult to imagine the variable of interest, p, being a missile system reliability. The requirement for missile reliability will usually be specified in a requirements document, and the decision maker will certainly be interested in the chances of the missile system meeting this requirement. If the requirement is for p to be at least 0.7, then using the cumulative distribution in Figure 4.2, he notes that the estimate of the probability that $p \geq 0.7$ is approximately 0.48. Since this probability is relatively small, the decision maker may also be interested in the fact that the estimate of the probability of exceeding 0.4 is 0.99. This additional information can provide valuable insight which may enable the decision maker to rationally specify a new acceptance criterion for p.

Another use of the posterior beta distribution of p occurs when using a Monte Carlo simulation to examine the uncertainty in some function of the variable p, where the function may or may not include elements of uncertainty other than p. Such a situation can be envisioned for the case previously considered where p represents a missile system reliability.

Suppose for example the variable of interest is the single shot kill probability;

$$P_{SSK} = R_{GSE} \cdot R_M \cdot P_{PF} \cdot M_L,$$

where

P_{SSK} = the single shot kill probability for the missile system,

R_{GSE} = the reliability of the ground support equipment,

R_M = the reliability of the missile,

P_{PF} = the probability of proper fuzing, and

M_L = the probability of a kill given proper fuzing.

The uncertainty in the single shot kill probability will depend on the uncertainty in the estimates of R_{GSE}, P_{PF}, and M_L as well as the estimate of R_M. The uncertainty in the estimate of R_M can easily

be introduced into a Monte Carlo simulation by sampling from the posterior cumulative distribution function of p.

According to the foregoing discussion, the posterior beta distribution appears to be a valuable decision-making tool. There are, however, certain other advantages and disadvantages which should be examined. The apparent artificial use of a beta distribution as a prior distribution in the updating procedure of Section 2 is certainly a questionable area. In relation to decision making, when a Bernoulli success parameter is the decision variable, several points can be made in defense of the beta distribution. First, the beta is of a form which lends itself quite readily to the distribution of a proportion. Its range is the unit interval; it is unimodal and can be skewed in either direction. Thus, by judicious choice of parameters, the beta probability density can easily be put into a form which will satisfactorily reflect one's prior beliefs. Since all available information concerning the parameter p should be used by the decision maker, the beta prior assumption has the additional advantage that it drastically simplifies the mathematics involved in the update procedure. Any last minute test results can readily be used to update the posterior beta distribution by merely repeating the update procedure with the posterior beta distribution now assuming the role of the prior beta distribution. Further, each update of the distribution reduces the impact of the subjectivity inherent in the initial prior assumption.

Several arguments against the Bayesian procedure also immediately come to mind. Foremost is the inherent subjectivity present in the technique. Certainly the classical analyst who firmly believes that the only legitimate types of probabilities emanate from frequency-of-occurrence data may find it difficult to accept the idea of using subjective or personalistic probabilities in forming a representative prior distribution. It is the Bayesian analyst's contention, however, that a reasonable decision maker will have intuition concerning an uncertain situation and will modify his feelings on the basis of sample or experimental evidence. He will certainly not blind himself to a

large portion of the information available merely on the basis that it may be subjective. As pointed out by Hamburg, "If only objective probabilities have meaning, then one cannot handle some of the most important uncertainties involved in problems of decision making."[4]

Another argument against the Bayesian procedure is that different analysts may come up with differing recommendations depending on their particular prior assumptions. In most situations, however, this argument is unwarranted since the individual assumptions are clearly visible and can be used as a basis for further arbitration.

On the positive side, a desirable feature of the subjective Bayesian approach is that although it allows the freedom to be conservative, it does not force conservatism on the analyst as do the classical techniques. The relevance of this point became evident in the comparison of the Bayesian and classical lower confidence limits in Section 3. Certainly a rational choice of a prior distribution representing the decision makers beliefs is far superior to the conservative classical viewpoint of ignoring a large portion of the available information. Drake summarized the Bayesian philosophy in the statement, "The Bayesian analyst believes that assisting in the consistent employment of all available data for a decision is part of his job, rather than a task to be left to some mysterious decision maker a few echelons up."[5]

Many arguments comparing the classical and Bayesian philosophy are available in other sources (e.g., References 6 and 7). It is not the purpose of this section to dwell on these philosophical implications.

---

[4]Hamburg, Morris, Statistical Analysis for Decision Making, Harcourt, Brace and World, Inc., New York, 1970.

[5]Drake, Alvin W., Bayesian Statistics for the Reliability Engineer, Proceedings of the National Symposium on Reliability and Quality Control, IEEE, 1966, pp. 315-320.

[6]Pozner, A.N., A New Reliability Assessment Technique, Technical Conference Transactions, American Society for Quality Control, 1966, pp. 188-201.

[7]Breipohl, A.M., R.R. Prarie, W.J. Zimmer, A Consideration of the Bayesian Approach in Reliability Evaluation, IEEE Transactions on Reliability, October, 1965, pp. 107-113.

It is, however, intended to demonstrate that the Bayesian philosophy does lend itself to decision problems concerning the Bernoulli parameter, p. In summary, the relevant points are:

- The fears concerning the Bayesian assumptions are often unwarranted.
- The decision maker can easily relate to statistical interpretations based on the posterior beta distribution.
- The Bayesian procedure should be given consideration in Bernoulli type decision problems.

One area of Bayesian analysis which needs further discussion is constructing the prior beta distribution. There are two basic problems to be considered in constructing the prior distribution:

- What is the prior distribution?
- Should the prior distribution dominate the posterior distribution?

Very simply stated, the prior will dominate the posterior distribution if the number of prior observations (m) exceeds the number of update observations (n). Recall that for this class of problems, it is assumed that some test data exist on which to base the prior distribution. These two problems depend on three basic considerations·

- How representative are the prior observations or is a significant difference in the update success proportion likely?
- How many prior and update observations are there?
- Does the prior beta distribution described by the specified parameters reasonably reflect the uncertainty in the estimate?

These three considerations which reflect the state of prior knowledge can be taken into account in a rational manner to construct a meaningful prior distribution. The next section is devoted to a discussion of a proposed method for constructing a prior distribution for this Bernoulli-type problem.

5. METHOD FOR CONSTRUCTING THE PRIOR DISTRIBUTION

5.1 Introduction.

As discussed in Section 4, one of the major criticisms of the Bayesian procedure is that the selection of the prior distribution is

arbitrary. Although this argument is valid in some instances, this is certainly not always the case. The authors believe that, for the class of problems being discussed in this paper, there is a rational and systematic way of constructing the prior distribution. It is emphasized, however, that the method presented in this section is a suggested approach and should not be construed as the only rational approach to the problem.

5.2 Method - General Discussion.

The method described in this section is intended to provide a rational systematic approach for analyzing one's state of knowledge, taking into account three basic considerations for constructing a prior distribution. These considerations are:

- Are the prior observations representative of the update observations, i.e., Is a significant difference expected in the update proportion of successes?
- Are the number of prior observations greater than the number of update observations?
- Does this prior distribution reflect the uncertainty in the estimate, or are the limits of the prior distribution reasonable?

Test observations are assumed to be the foundation for constructing the prior distribution for the class of problems addressed in this report. For instance, if the success proportion of interest is production missile reliability, then the prior observations may be based on industrial prototype and R&D missile flights.

If design problems are diagnosed and design changes implemented, there may be some question as to just how representative these observations are of the population of interest. One potential aid in reducing the bias associated with design changes is to "no-test"* the design failures. Of course, even using this type of scoring criterion, there still may be reason to suspect that a significant improvement in the update

*A no-test is a test observation that has been eliminated from the data set.

observations is likely, since in all instances it is assumed that the update observations are from the population of interest.

As mentioned previously, these prior test observations serve as the foundation for constructing the prior distribution. For example, if there are ℓ successes out of m prior observations, then a beta distribution with parameters ℓ and $m-\ell$ would serve as the initial beta distribution.

This distribution is then modified by use of all available subjective information (based on engineering judgment, experience with similar systems, etc.) to form what will be referred to as the prior distribution.

It should be recalled that the two basic questions being addressed are: (1) What is the prior distribution? and (2) Should the prior distribution dominate the posterior distribution? In light of the initial beta distribution, these two questions can be addressed by examining the three basic considerations mentioned earlier.

Perhaps the most important consideration is the first one concerning whether the prior observations are representative of the update observations since it impacts both questions. Note that the first question (What is the prior distribution?) has two parts. The first part is whether the most likely value of the success proportion seems reasonable in light of all other prior information (i.e., Is a significant difference in the update success proportion expected?) The second part addresses whether the distribution accurately reflects the uncertainty. However, before the uncertainty consideration can be evaluated, one must determine if the dominance problem need be considered. This dominance problem depends on the first consideration. If the most likely value of the initial beta distribution does seem reasonable (i.e., no significant difference is expected), then whether the prior distribution dominates the posterior distribution doesn't really matter. In essence, one is indicating that the prior data are thought to be reasonable. Hence, the relative number of prior and update observations

is not important. The only important consideration remaining is whether a prior distribution with parameters ℓ and $m-\ell$ accurately reflects the uncertainty in the estimate of the success proportion.

On the other hand, if there is reason to suspect a significant difference in the most likely value of the success proportion, then one would want to shift the prior distribution to a more reasonable value. Of course, what constitutes a significant difference in the most likely value is dependent on both the specific problem and the analyst's judgment.

The problem now confronting the analyst is how can this prior distribution be shifted? One way of doing this is to decrease the number of observations (m) by some number, e, while keeping the number of successes (ℓ) constant. This is a reasonable approach when the most likely value (ML) is thought to be low; e can be obtained algebraically in the following manner. Simply specify a more reasonable most likely value and solve the following equation for e:

$$ML^* = \frac{\ell-1}{m-e-2} .$$

This is simple to apply since ℓ and m are known and it is intuitively appealing since it precludes assigning more weight to a prior distribution than the available data would suggest.

On the other hand, if the most likely value is thought to be high, then one could reduce the number of successes (ℓ) by some fixed number, e, while keeping the number of observations constant. Once again one can solve for e by specifying a more reasonable most likely value and solving the equation

$$ML = \frac{\ell-e-1}{m-2} .$$

Note that this latter case is not thought to be very likely in realistic situations, but it is presented for completeness sake.

*The most likely value of a beta distribution is $ML = \frac{\ell-1}{m-2}$.

After determining e,* a new beta distribution exists with parameters ℓ and $m-e-\ell$ or $\ell-e$ and $m-\ell+e$, depending on the particular situation. Given either of these sets of parameters, the problem of whether the prior distribution should dominate the posterior distribution becomes important. To examine this, one must consider whether the number of prior observations (i.e., m or m-e) is greater than the number of update observations, since the distribution having the most observations will have the greatest impact on the posterior distribution. Recall from Section 2 that the posterior mean is the weighted average of the prior and update success ratio. If one suspects a significant difference in the success proportion, then the update observations should have at least equal weight or dominate the posterior distribution.

Whether the update observations should dominate the posterior distribution depends on whether this prior distribution reflects the uncertainty in the estimate of the success proportion. The next problem for the analyst is: How does one determine whether the prior accurately reflects the uncertainty in the estimate? Generally, a few brief calculations and a plot** should provide all of the information needed to evaluate whether the prior accurately reflects the uncertainty. For instance, if the limits of the prior distribution are 0.4 to 0.6 (i.e., 99 percent the area of the distribution is contained within these limits), but one suspects that the success proportion can take on values between 0.6 and 0.75, then this distribution does not accurately reflect the uncertainty in the estimate.

The final problem confronting the analyst is how does one make this prior more accurately reflect his state of knowledge. This is achieved in general by reducing the weight of the prior. For instance, if the prior data were 10 successes out of 20 observations, then this could be treated as 5 successes out of 10 observations or 2 successes out of 4 observations by assigning a 0.5 or 0.2 weight*** to the prior observations.

*Note e should be rounded off to the nearest integer.

**This will be explained in detail in the example.

***Throughout this report W is used as the symbol for prior weight.

Selecting a prior weight is not an exact science and should not be approached as such. The best approach is to reduce the prior weight by 0.25 or 0.2, examine this new prior distribution, and decide if it more accurately reflects the uncertainty in the estimate. If it does, then stop. If it doesn't, then repeat the procedure with a smaller prior weight. Each time reduce the prior weight by the same amount.

Before describing the method in detail, some state vector notation must be introduced. Since there are three basic considerations, all of the states of knowledge can be described with a three-dimensional state vector. In addition, all of the considerations can be handled with yes/no logic (i.e., the dimension of each component of the vector is two); therefore, there are eight possible states of knowledge (vectors). Thus, the state vector S(NO, YES, YES) represents a state in which the responses to the three basic considerations a, b, and c are NO, YES, and YES respectively.

Finally, two points should be made. First, the previously mentioned considerations are not independent. For instance, if one expects a significant difference in the update proportion of successes, it is likely that the prior observations will not reflect the uncertainty. In addition, if the number of prior observations is greater than the number of update observations, one would probably want to reduce the prior weight (i.e., the prior should not dominate the posterior distribution in this case). This method provides a framework for analyzing these three considerations sequentially. After each has been analyzed separately and the initial beta parameters modified in light of the first consideration, the total state of knowledge can be evaluated, and the trade-offs considered in weighting the prior distribution. A more detailed discussion of the trade-offs is deferred until the method is described in detail.

Second, the range of the prior weight is restricted to a number greater than zero and less than or equal to one (i.e., $0 < W \leq 1$).

The reason for the lower bound is obvious. It is not meaningful to talk about negative weights. Further, it is assumed that there is some useful information in the prior observations. Thus, a prior weight of zero, indicating no useful information, is not considered. The rationale for the upper bound warrants a more detailed explanation. To assign a weight greater than one would imply more certainty than the data reflect. Even if the observations for the prior were taken from the same lot, one would not want to count each observation more than once. For example, if the number of defective items in a lot are being estimated and two samples are drawn from this lot, then there is no rational basis for counting observations from the one sample more than the observations from the other. Therefore, in situations where the prior observations are not from the same lot, it is not reasonable to assign a weight greater than one.

5.3 Method.

Recalling the three primary considerations, Figure 5.1 depicts the eight possible states of knowledge. Starting at the top of the flow diagram, the user begins by asking the question "Is there any reason to suspect a significant difference in the update proportion of success?" If the answer to this question is no, then the analyst must consider whether there are more prior observations than update observations. If there are fewer prior observations, then the next stage is to answer the question, "Does this prior reflect the uncertainty in the estimate?" As mentioned previously, this can be done by examining the prior distribution (i.e., the range and standard deviation). If the answer to this question is no then the state vector S(NO, NO, NO) has been obtained. Hence, the flow diagram provides a systematic questioning procedure for determining the state vector, modifying the prior distribution in light of the first consideration, and assigning a weight to the prior distribution. For each of the eight possible state vectors, guidance is given for assigning a prior weighting factor. In some instances, the guidance is specific while in others it provides an

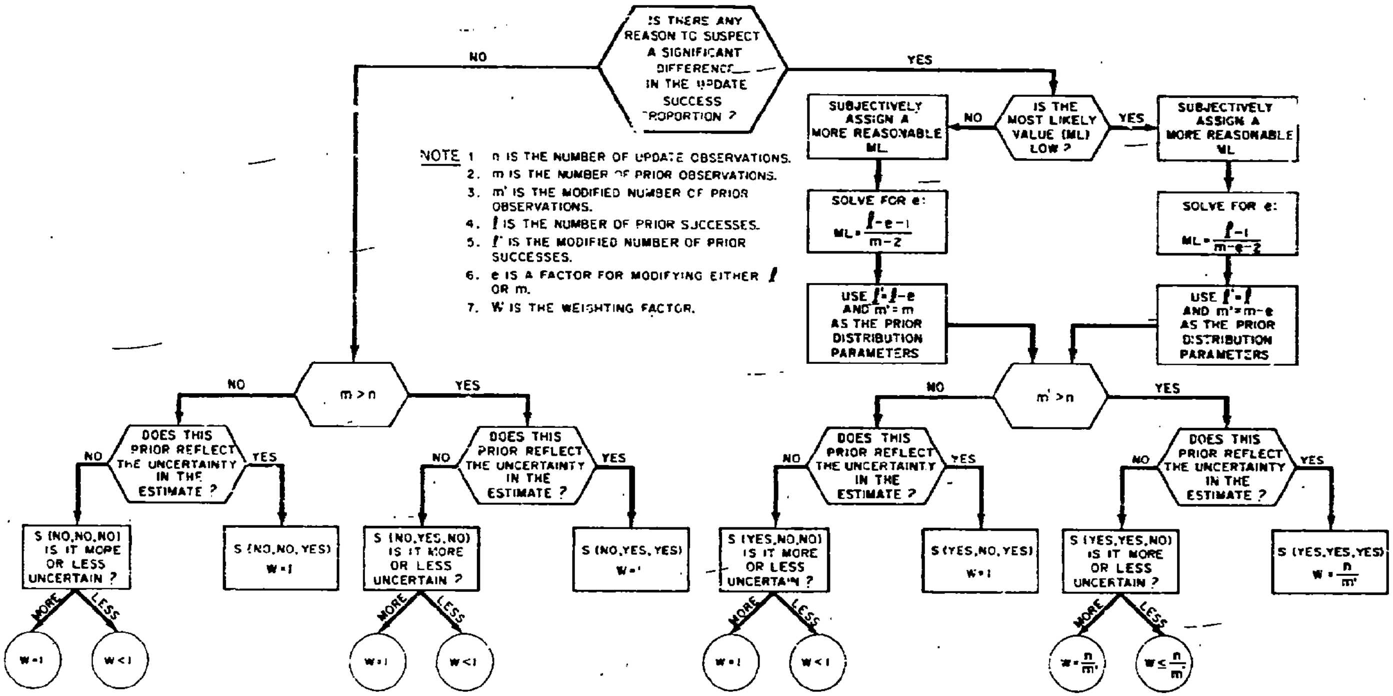

Figure 5.1 Method for Constructing the Prior Distribution

upper bound. In the next few paragraphs, each of the possible state vectors will be analyzed, and the rationale for recommending a particular weighting factor will be discussed.

Going from left to right, the components of the state vector will correspond to considerations a, b, and c, respectively. (See Figure 5.1.)

a. S(NO, NO, NO) is the state vector indicating that there is no reason to suspect a significant difference in the success proportion; the number of prior observations is less than or equal to the number of update observations and, based on available information, the prior does not reflect properly the uncertainty in the estimate. Given this state vector, one must now ask, "Is the prior distribution more or less uncertain than is thought to be reasonable?" A prior distribution is said to be more uncertain if the range of the distribution is greater than is thought to be reasonable. In this case, a prior weight equal to one should be used (i.e., if there are ten observations, then count them as ten observations). The rationale for using a prior weight of one is:

- There is no reason to suspect a drastic difference in the update success proportion; hence, there is no reason to adjust the most likely value of the success proportion.
- Since the number of prior observations is less than or equal to the number of update observations, the prior will not dominate the posterior distribution.
- The prior is thought to be more uncertain than would appear reasonable. The way to decrease the uncertainty is to increase W (i.e., W > 1), but to do this would exceed the upper bound on W.

On the other hand, if for S(NO, NO, NO) the prior estimate is thought to be less uncertain* than is reasonable, then a weight less

*This is perhaps a bad choice of words since less uncertain really implies more certainty in the estimate.

than one, which will in effect cause the uncertainty to be reflected more reasonably, should be used. The exact value of W depends on the problem.

b. S(NO, NO, YES) is the same state vector as S(NO, NO, NO) except that the prior distribution does reflect the uncertainty in the estimate. Thus, it is reasonable to assume that the rationale for the first two elements of the state vector is the same (i.e., no adjustments to the prior weight or most likely value). Because the prior distribution does reflect the uncertainty in the estimate, there is no reason to adjust the prior weight. Therefore, a prior weight of one should be used in this case.

c. S(NO, YES, NO) is the state vector which indicates that there is no reason to suspect a significant difference in the update success proportion, that the number of prior observations is greater than the number of update observations, and that the prior does not reflect properly the uncertainty in the estimate. Once again the analyst is faced with two possibilities. Is the prior distribution more or less uncertain? If it is more uncertain, a prior weight greater than one would have to be used, in effect, to decrease the uncertainty. However, this is not justified since a weight greater than one exceeds the upper bound on W. Therefore, a prior weight of one is recommended.

On the other hand, if the prior distribution is less uncertain, the prior weight should be less than one. How much less than one depends on the particular problem and the amount by which it is felt the prior fails to reflect properly the uncertainty. Once again there was no reason to adjust the most likely value of the prior distribution.

d. S(NO, YES, YES) is the same state vector as S(NO, YES, NO) except that the prior distribution does reflect properly the uncertainty in the estimate. Based on this state of knowledge, there is no reason to adjust the prior weight (i.e., W=1) or the most likely value. Of course, some analysts might argue that the prior distribution should

never dominate the posterior distribution, but this is not thought to be valid in light of the state of knowledge. However, if the analyst feels strongly about this, the weight could be reduced, but the lower limit should be $W = \frac{n}{m}$.* This weight, W, would give the prior and update equal weight in the posterior distribution.

Before continuing with a discussion of the selection of a prior weight for the remaining states, it should be pointed out that in all of the remaining states the initial beta distributions will be shifted to more reasonably reflect the most likely value of the success proportion. All of the guidance given for the prior weight will then apply to the modified beta distribution.

e. S(YES, NO, NO) is the state vector which indicates that there is reason to suspect a significant difference in the update success proportion, that the number of prior observations is less than or equal to the number of update observations, and that the prior distribution does not reflect the uncertainty in the estimate. Before continuing, one must ask the following question "Is the prior distribution more or less uncertain than is thought to be reasonable?" If it is more uncertain, use a prior weight of one. The rationale for this is:

- There is a reason to suspect a significant difference in the update success proportion; hence one would probably want the update to have at least as much or more weight in the posterior distribution. Therefore, the weighting factor selection is dependent on the number of prior and update observations.
- For this state vector, the number of prior observations is less than or equal to the number of update observations. Hence, the update information will have at least equal weight (even if W=1).
- Finally this prior distribution is thought to be more uncertain; therefore, this would tend to suggest a weight greater than one. Once again this is unrealistic, because a prior weight greater than one exceeds the upper bound on W. In addition, the fact that there is reason to suspect a

*m is the number of update observations and n is the number of prior observations.

drastic difference in the success proportion of the update observations would tend to imply weighting the prior distribution less. In this instance, the update observations will have at least equal weight, and it is not necessary to reduce the prior weight. Therefore, in light of all of this information, a prior weight of one is thought to be most reasonable.

On the other hand, if the prior distribution is less uncertain than is thought to be reasonable, the prior weight should be less than one. Again the exact value is a function of the particular problem.

f. S(YES, NO, YES) is the same state vector as S(YES, NO, NO) except that the prior distribution does reflect properly the uncertainty in the estimate. Based on all of this information and the rationale for the first two components of S(YES, NO, NO) a prior weight of one is recommended.

g. S(YES, YES, NO) is the state vector that indicates that there is a reason to suspect a significant difference in the success proportion, the number of prior observations is greater than the number of update observations and the prior distribution does not reflect properly the uncertainty in the estimate. Again the analyst must ask the question, "Is the prior distribution more or less uncertain than is thought to be reasonable?" If it is more uncertain then use a prior weight $W = \frac{n}{m'}$. The reasons for selecting this weight are as follows:

- A drastic difference is suspected in the update success proportion. Therefore, one would probably want the update to have at least as much weight as the prior, and a prior weight of $W = \frac{n}{m'}$ satisfies this requirement.
- However, for this state vector the number of prior observations is greater than the number of update observations. Using a prior weight of one, the prior would dominate the posterior. Therefore, it seems reasonable to use as the greatest prior weight $\frac{n}{m'}$. The question that still remains is should the prior weight be less than $W = \frac{n}{m'}$.

- Since the prior distribution is more uncertain than is thought to be reasonable, one might be inclined to use a prior weight greater than one, but once again a prior weight greater than one exceeds the upper bound on W. Because the prior distribution should not dominate the posterior distribution and the prior distribution is already thought to be more uncertain than is reasonable, a prior weight, $W = \frac{n}{m'}$, is thought to be the best compromise.

  This trade-off accepts a little more uncertainty in the prior estimate while allowing equal weight to be given the update distribution.

On the other hand, if the prior is less uncertain than is thought to be reasonable the prior weight should be less than or equal to $W = \frac{n}{m'}$. The reason is that even if a weight greater than $\frac{n}{m'}$ would reflect the uncertainty in the estimate, one would still want the update observations to have at least equal weight in the posterior distribution.

h. S(YES, YES, YES) is the same state vector as S(YES, YES, NO) except that the prior distribution does reflect the uncertainty in the estimate that is thought to be reasonable. Therefore, a prior weight of $W = \frac{n}{m'}$ is recommended. The trade-off again is in terms of having the update weigh as much as the prior and increasing the uncertainty in the estimate. Once again it is thought that the update should have least equal weight, but this is constrained by the increase in the uncertainty. Therefore, it is thought that a prior weight of $W = \frac{n}{m'}$ is the best compromise.

Before continuing with an example, it should be emphasized again that this method is a general framework for systematically shifting the prior distribution and selecting a prior weight in light of the analyst's state of knowledge. This is not a prior weight index; other weights might be assigned equally well under the same logic. Its application is indeed largely a matter of personal preference and intuition. As indicated earlier, it is always good procedure to test the sensitivity of the posterior solution to the prior weight selected. The example that follows should give some insight into the practical application of this method.

6. EXAMPLE

6.1 Background.

To illustrate the application of the Bayesian procedure and the method for constructing the prior distribution, the following hypothetical decision problem is described. Assume that the US Army is developing a surface-to-air missile to provide forward air defense for the Field Army. The tactical production decision is to be made in about a year, and to date there have been test firings with Research and Development rounds (40 firings) and Industrial Prototype rounds (30 firings). In the near future, the Initial Production Tests are to be initiated, and it is anticipated that by the decision date there will be 20 test firings with production missiles. One of the important questions facing the decision maker is, "Will the system meet the production missile reliability (R_M) requirement?"

Unfortunately, only a limited amount of production test flight data will be available by the decision date, and if only production missile test data are used to estimate R_M, then a great deal of potentially useful information is being ignored. Further compounding the problem is the fact that the contractor is claiming that the quality control program at the manufacturing plant has been improved, and as a consequence the reliability is significantly higher than that demonstrated to date by non-production rounds. The contractor's past performance and the fact that no concrete procedure changes have been instituted at the manufacturing plant make one suspect the claim. Therefore, the problem is how can the non-production missile data and all other pertinent information be meaningfully combined with the production data for decision-making purposes.

6.2 Scoring of Missile Flights.

The results of a hypothetical scoring of the Research and Development, Industrial Prototype and Production missiles are summarized in Table 6.1. Developing a rationale for scoring non-production and

production flights is a large task, but not an impossible one. No example of a scoring criterion is provided here because it is not thought to be germane to this example. One point that should be made, however, is that the objective in developing a scoring criterion should be to remove all possible biases. For instance, if as a result of the non-production flights, design problems were diagnosed and corrected then these flights should not be counted as observations. Based on the hypothetical scoring of the missile test flights in Table 6.1, there are 40 observations for the pre-production rounds and 20 observations for production rounds (i.e., "no tests" do not count as observations).

TABLE 6.1 MISSILE FLIGHT FIRING SUMMARY

Type of Missile	Successes	Failures	No Tests
Research and Development	15	10	15
Industrial Prototype	10	5	15
Production	16	4	0

6.3 Application of the Method for Constructing the Prior Distribution.

To apply the Bayesian procedure described in this report in a real-life situation, the problem must have the attributes described earlier, (Bernoulli process test data for update and prior) which this problem obviously does.

When we recall the method which was presented in Section 5 for constructing the prior distribution, we ask the question, "Is there any reason to suspect a significant difference in the success ratio for production missiles?" In this hypothetical example, this ratio, based on contractor claims and on development and test agency engineering judgement, is suspect.

All of the expert judgment indicates that the most likely value of the initial beta distribution is low. Therefore, a more reasonable value must be specified and $ML = \frac{\ell-1}{m-e-2}$ must be solved for e.

From the information obtained by the development agency engineers, contractor engineers, and test agency engineers, a more reasonable most likely value is thought to be 0.75. Given ML = 0.75, ℓ=25, and m=40, the previous equation can be solved for e. In this instance e=6, and the adjusted prior beta distribution now has parameters, 25 and 9.

Next, there are 34 prior observations and 20 update observations. In addition, the prior is less uncertain than is thought to be reasonable. Hence the state vector is S(YES, YES, YES), and the prior weight recommended is $W \leq \frac{n}{m'}$ (i.e., $W \leq \frac{20}{34}$). Since 20/34 is approximately 0.59, an upper bound of 0.6 would probably be used for computational ease. The reason for this bound on the prior weight is that, since a significant difference in the update success proportion is expected, the update observations should have at least as much weight as the prior observations in the posterior distribution. The exact value of W depends on which prior weight less than or equal to 0.6 will most reasonably reflect the uncertainty in the estimate.

To illustrate the impact of a dominant prior consider the following example. Suppose a significant difference in the update success proportion was expected, and after shifting the most likely value of the prior distribution, there were 16 successes out of 40 observations. Even though an effort has been made to select a reasonable prior distribution, one would still want the update information to have at least as strong an influence as the prior information in the posterior distribution. If a prior weight of one is used, and there were 13 successes in 20 update observations (i.e., a 0.65 success proportion), then this extreme variation in the update success proportion would not be adequately represented in the posterior distribution. In this instance, the mean of the posterior distribution of the reliability is equal to 0.483. However, the probability of the true reliability exceeding 0.65 (the success proportion for the update observations) is almost zero, which is not reasonable if the update information is to be emphasized. A

maximum prior weight of 1/2 will give at least equal weight to the update information and allow the update to have at least an equal influence on the posterior distribution.

Before continuing, two points should be emphasized. The first is that the determination of the prior weight is not an exact quantitative science and should not be approached as such. The question which is to be answered is whether the prior should or should not dominate the posterior. In the preceding example, distinguishing between prior weights of 0.45, 0.50, or 0.55 is meaningless. It is probably not possible to make such a fine distinction. The second consideration is that another set of circumstances could yield an entirely different prior weight.

Using a prior weight equal to 0.6, the parameters of the prior beta distribution for the missile example are 15 and 5.4, respectively. These are based on 25 successes in 34 pre-production missile test firings. For computational ease, the parameters can be rounded off to 15 and 5 without any significant impact on the final results.

The last step in applying the method is for the analyst to decide if the prior distribution should be weighted less than 0.6. Weighting the prior distribution less than 0.6 depends on whether a prior weighted by 0.6 properly reflects the uncertainty.

In the missile example, the prior distribution with $W = 0.6$ has as its mean 0.75 (mode 0.78) and the limits of the distribution are approximately 0.46 and 0.98 (see Figure 6.1). The question is, "Is the true estimate likely to lie outside the limits of the distribution?" In this example, the limits of the distribution are thought to be reasonable. On the other hand, if the limits of the distribution are significantly narrower than is thought to be reasonable, then the prior weight can be further reduced to reflect this uncertainty. It should be noted that reducing the prior parameters by some factor merely increases the variance or spread of the prior distribution but does not affect the mean.

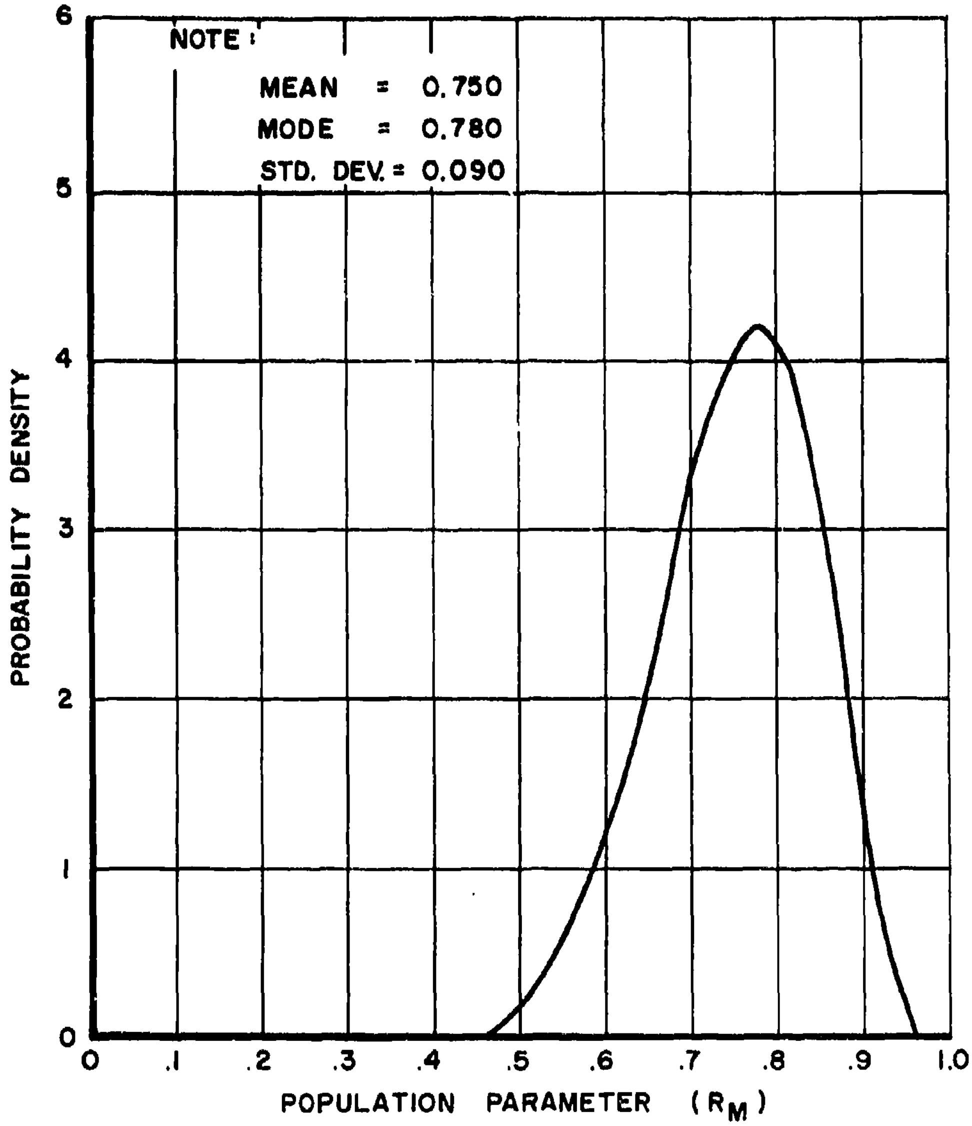

Figure 6.1 – Beta Prior Probability Density Function

Given the parameters of the prior distribution (ℓ=15 and m-ℓ=5) and the update distribution parameters (k=16 and n-k=4) in this example, the posterior probability density function of R_M is

$$f_{R_M|16}(R_M|16) = \frac{\Gamma(40) R_M^{30}(1-R_M)^8}{\Gamma(31)\Gamma(9)} \quad 0 < R_M < 1.$$

This distribution has as its mean 0.775 and standard deviation 0.07 (see Figure 6.2). If, in this hypothetical example, the R_M requirement is 0.85 or greater, then by use of the foregoing posterior probability density, the probability of achieving this requirement is approximately 0.12 (see Figure 6.3).

While this is not a favorable result, the following steps can be taken:

- Some less stringent requirement could be evaluated (e.g., the probability that R_M is greater than or equal to 0.7).
- The distribution could be examined to determine the lower limit.
- One could examine the sensitivity of the prior distribution.

However, the sensitivity analysis should not be conducted indiscriminately (i.e., don't play a numbers game). There should be a legitimate reason for changing the prior distribution. These reasons will generally revolve around debate over the rationale (assumptions) for selecting the most likely value and/or the prior weight. For example, there may be two distinct opinions about the prior weight; one group may be optimistic (smaller weight) while the other group may be pessimistic (larger weight). After analyzing the rationale behind both of these opinions, the analyst may have selected a weight somewhere between these two schools of thought. In this example it is legitimate to do some sensitivity analysis to examine the impact of the optimistic and/or pessimistic point of view.

The only one of these three possible activities that deserves illustration is sensitivity analysis. In this example, only the prior

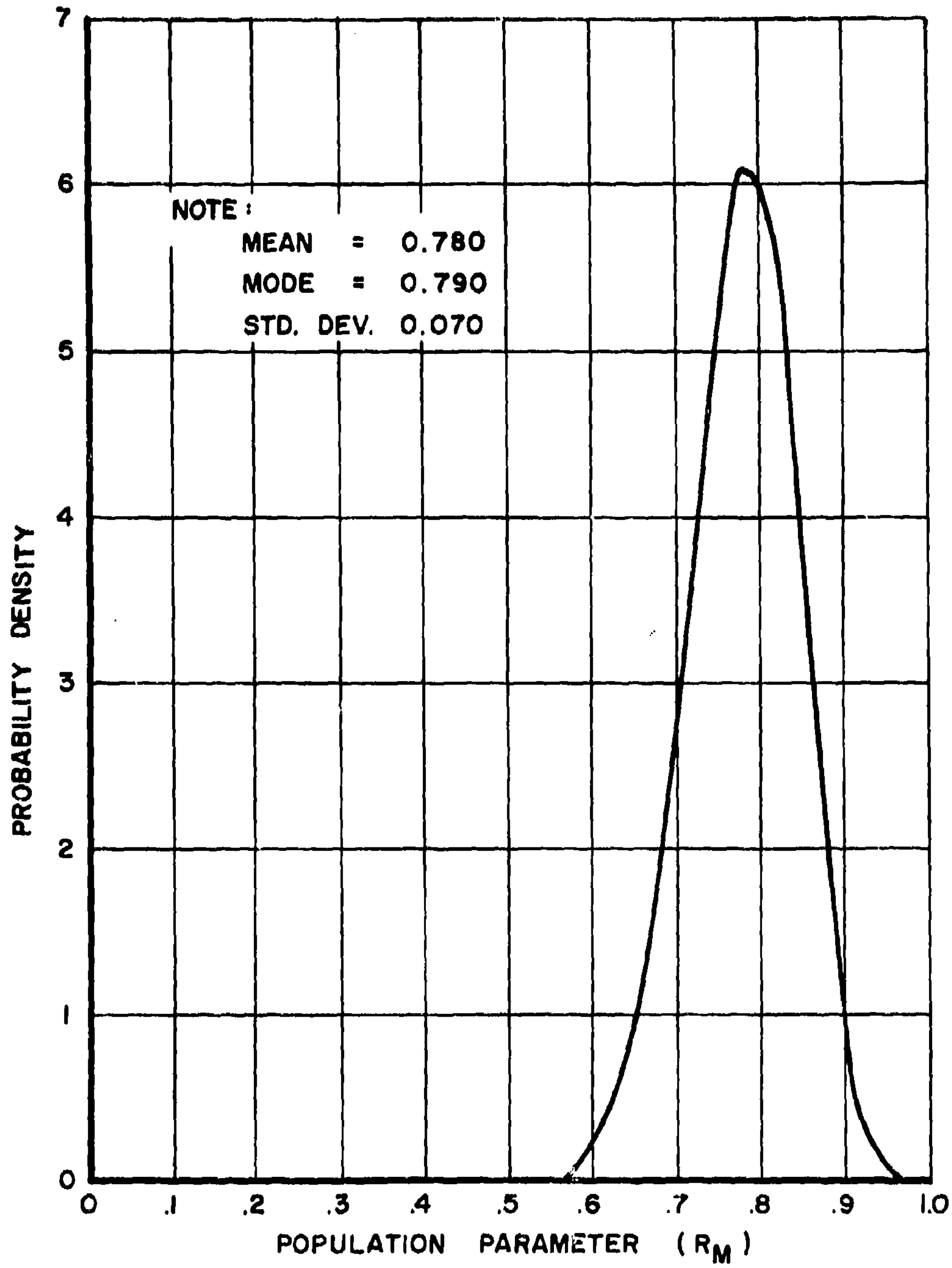

Figure 6.2 – Beta Posterior Probability Density Function

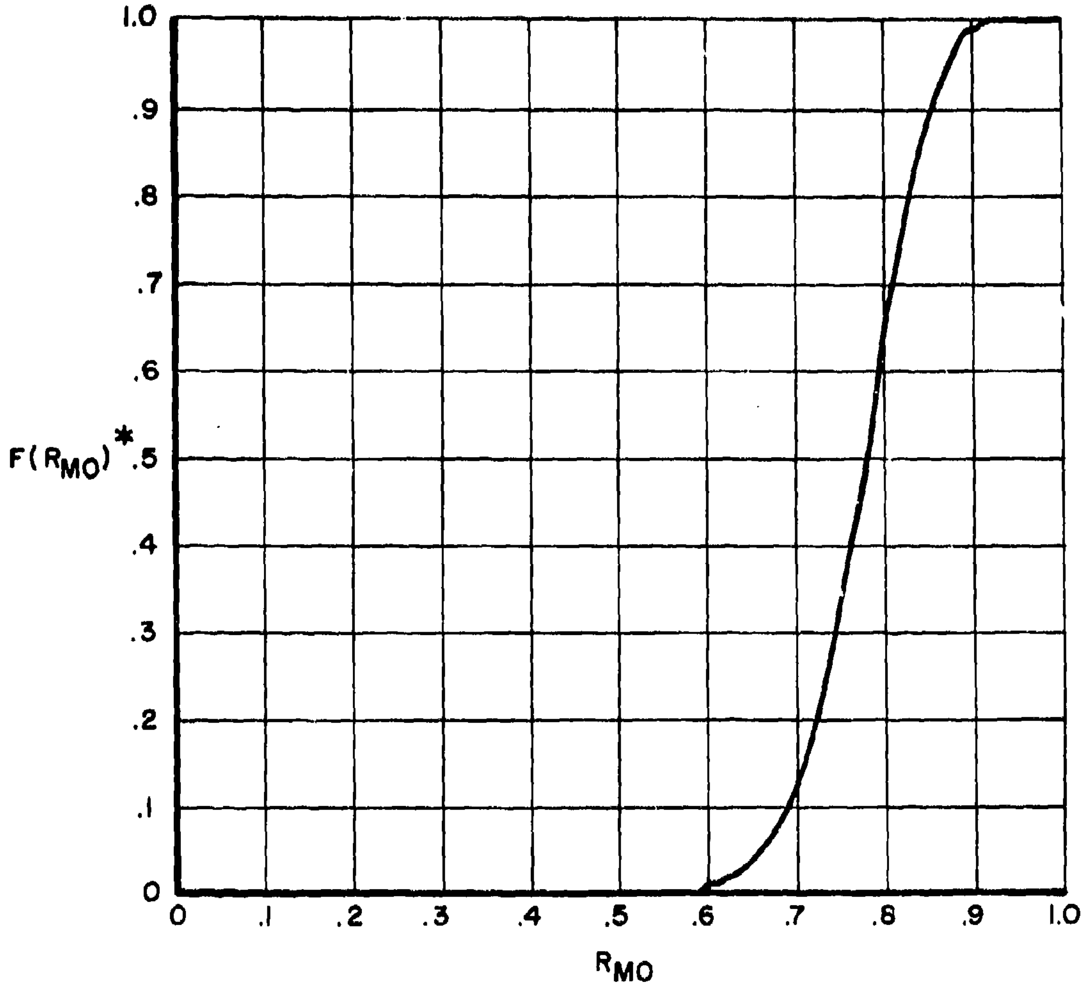

Figure 6.3–Cumulative Posterior Beta Distribution

* $F(R_{MO}) = P\left[R_M \leq R_{MO}\right]$ (i.e., THE PROBABILITY THAT THE TRUE RELIABILITY (R_M) IS LESS THAN OR EQUAL TO R_{MO})

weight will be modified. To examine the impact of being optimistic, a prior weight of 0.4 is used. By use of this weight, the parameters of the prior distribution are 10 and 4, respectively. This gives rise to the following posterior probability density function:

$$f_{R_M|16}(R_M|16) = \frac{\Gamma(34) R_M^{25}(1-R_M)^7}{\Gamma(26)\Gamma(8)} \quad 0 < R_M < 1,$$

with mean 0.76 and standard deviation 0.07 (see Figure 6.4). Based on this new posterior distribution the probability that R_M is greater than or equal to 0.85 is now 0.11 (see Figure 6.5). Hence, the posterior in this case is not sensitive to a prior weight change of 0.2.

What is really being done in the sensitivity analysis is that the uncertainty in the estimate is being increased or decreased as the number of prior observations decreases or increases while the mean and mode are being shifted toward or away from the update success ratio.

The preceding example serves to illustrate the Bayesian procedure and the method for constructing a prior weight. It should also demonstrate that one can systematically evaluate and combine relevant objective and subjective information for decision making purposes. The Bayesian procedure described in this paper was used to estimate the uncertainty in the estimate of missile reliability in a recent study with little more effort then is normally required for a reliability evaluation using the classical procedures. This estimate was then used in a Monte Carlo simulation to estimate the distribution of effectiveness for the missile against the postulated threat in the various modes of attack. Based on this application, the procedure was found to be of significant value for analysis in support of the decision making process. The need for having a systematic procedure for analyzing one's state of knowledge became apparent in the application, and the method described in Section 5 was developed.

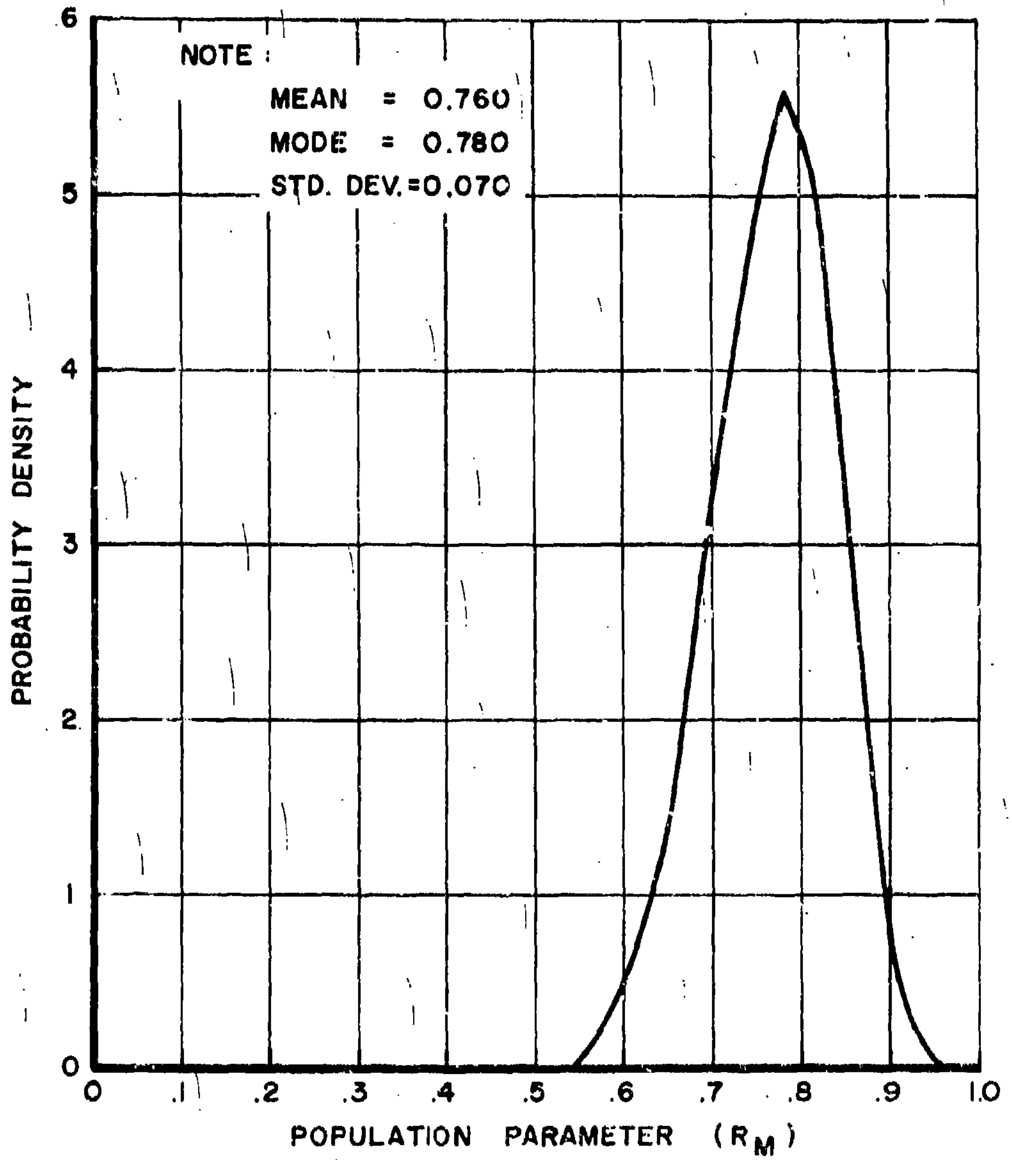

Figure 6.4 - Beta Posterior Probability Density Function

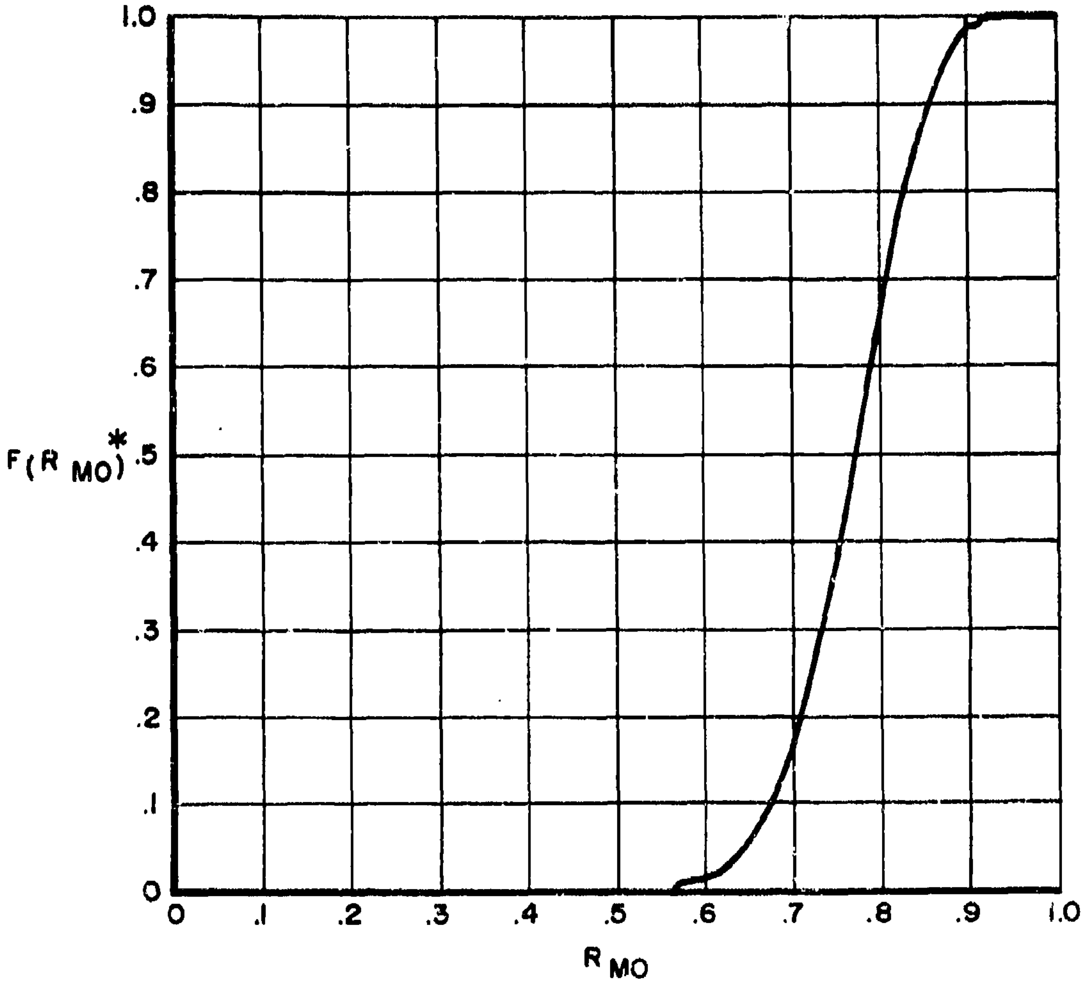

Figure 6.5 – Cumulative Posterior Beta Distribution

* $F(R_{MO}) = P\left[R_M \leq R_{MO}\right]$ (i.e., THE PROBABILITY THAT THE TRUE RELIABILITY (R_M) IS LESS THAN OR EQUAL TO R_{MO}.)

7. SUMMARY AND CONCLUSIONS

This report is concerned with decision making under uncertainty for the class of problems where the decision variable is the Bernoulli success probability, p. The problem is analyzed from classical and from Bayesian points of view.

Historically, either classical or Bayesian point or interval estimation has been the standard approach to this problem. In using classical techniques, it is difficult to account for all of the information concerning the unknown quantity which comes from any source other than the particular sample which has been taken. Further, none of the above approaches, either classical or Bayesian, addresses the decision problem directly. By use of the results of these procedures, one cannot make probability statements about meeting or exceeding a specific requirement for p, nor can one readily examine the uncertainty in p for the purpose of defining a more reasonable requirement for p.

As discussed in Section 4, the use of the posterior beta distribution, obtained in Bayesian updating, is a viable alternative to the above-mentioned procedures. It takes into account prior information and can also be used directly in the decision-making process.

Unfortunately, there still seems to be some mystique surrounding any application of Bayesian statistics. This is due in some instances to a disagreement with the Bayesian philosophy and in others to the lack of a true understanding of the mechanism of the Bayesian approach. In this respect, many of the popular objections have been examined and found to be unwarranted for this class of problems. Perhaps one of the most widely used arguments against the use of the Bayesian procedure is the apparent absence of a rational basis for constructing a prior distribution. For this class of problems, however, the argument has very little substance since, in general, there will certainly be a basis for selecting the form of the prior distribution, and there does exist a rationale basis for constructing a prior distribution, as evidenced by the suggested method in Section 5.

In relation to point and interval estimation, Section 3 contains a detailed comparison of the classical maximum likelihood and Bayesian point estimates with respect to expected squared error loss. It also contains a comparison of the lower confidence limits resulting from a classical and a Bayesian approach. In both these instances it is shown that, in many non-trivial practical situations, the Bayesian procedures provide more realistic estimates when using minimum expected squared error loss and greatest lower bound as the criterion for determining the best point and interval estimates.

Thus, it is the contention of the authors that the Bayesian approach, although not to be applied indiscriminately, should be given serious consideration when drawing inferences concerning the Bernoulli process success probability, p. This is particularly true in the decision making context.

REFERENCES

1. Mood, A.M., F.A. Graybill, Introduction to the Theory of Statistics, Second Edition, McGraw-Hill Book Company, New York, 1963.

2. Pearson, K. Ed., Tables of the Incomplete Beta - Function, University Press, Cambridge, England, 1934.

3. Benton, Alan W., An Investigation of the Characteristics of Bayesian Confidence Intervals for Attribute Data; Technical Memorandum No. 14, November 1969; Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland.

4. Hamburg, Morris, Statistical Analysis for Decision Making, Harcourt, Brace and World, Inc., New York 1970.

5. Drake, Alvin W., Bayesian Statistics for the Reliability Engineer, Proceedings of the National Symposium on Reliability and Quality Control, IEEE, 1966, pp. 315-320.

6. Pozner, A.N., A New Reliability Assessment Technique, Technical Conference Transactions, American Society for Quality Control, 1966, pp. 188-201.

7. Breipohl, A.M., R.R. Prarie, W.J. Zimmer, A Consideration of the Bayesian Approach in Reliability Evaluation, IEEE Transactions on Reliability, October 1965, pp. 107-113.

Next page is blank.

APPENDIX

BETA TO F TRANSFORMATION

A random variable U is said to have a beta distribution with parameters a and b, if its probability density function is of the form

$$f(u;\ a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, u^{a-1}\ (1-u)^{b-1}$$

$$0 < u < 1$$
$$a > 1$$
$$b > 1.$$

Consider the transformation

$$U = \frac{\frac{a}{b}\,V}{1+\frac{a}{b}\,V} \tag{1}$$

The Jacobian of the transformation (1) is

$$J = \frac{dU}{dV} = \frac{\frac{a}{b}}{(1+\frac{a}{b})\ V^2}$$

Thus, the probability density function of V is given by

$$g(v;\ a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left[\frac{\frac{a}{b}v}{1+\frac{a}{b}v}\right]^{a-1}\left[1 - \frac{\frac{a}{b}v}{1+\frac{a}{b}v}\right]^{b-1}\ \frac{\frac{a}{b}}{(1+\frac{a}{b}v)^2}$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\ \left(\frac{2a}{2b}\right)^{a}\ v^{a-1}\ (1+\frac{a}{b}v)^{-(a+b)}$$

$$0 < v < \infty$$
$$a > 1$$
$$b > 1$$

which is the probability density function of a random variable having an F distribution with 2a and 2b degrees of freedom.

Using transformation (1), one can make probability statements concerning a beta variate using tables of the F distribution (This is desirable since they are more available.). For example the 100(1-a) percent lower confidence limit for the Bernoulli success probability is given by the value of P_L which satisfies the equation

$$P[p_L < U < 1] = 1-\alpha \tag{2}$$

where U has a beta distribution with parameters a and b. Using transformation (1) probability statement (2) is seen to be equivalent to

$$P[p_L < \frac{\frac{a}{b} V}{1+\frac{a}{b} V} < 1] = 1-\alpha \tag{3}$$

where V has an F distribution with 2a and 2b degrees of freedom.

After some algebraic manipulation, statement (3) is seen to be equivalent to

$$P[0 < V' < \frac{a}{b} (\frac{1}{p_L} - 1)] = 1-\alpha \tag{4}$$

where V' has an F distribution with 2b and 2a degrees of freedom. The solution to equation (4) is then given by

$$p_L = \frac{1}{1+\frac{b}{a} v'_{1-\alpha}}$$

where $v'_{1-\alpha}$ is the $100(1-\alpha)$ percent point of the F distribution with 2b and 2a degrees of freedom.